AD_____

GRANT NUMBER DAMD17-94-J-4174

TITLE: Utilizing Serial Measures of Breast Cancer Risk Factors

PRINCIPAL INVESTIGATOR: Mimi Y. Kim, Sc.D.

CONTRACTING ORGANIZATION: New York University Medical Center
New York, New York 10010-2598

REPORT DATE: February 1998

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1998	Annual (31 Dec 96 - 30 Dec 97)

4. TITLE AND SUBTITLE
Utilizing Serial Measures of Breast Cancer Risk Factors

5. FUNDING NUMBERS
DAMD17-94-J-4174

6. AUTHOR(S)
Mimi Y. Kim, Sc.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
New York University Medical Center
New York, New York 10010-2598

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200*

A model for adjusting hormone measurements for systematic fluctuations over the menstrual cycle, which will allow more valid comparisons of hormone levels between pre-menopausal cases and controls, was developed. The model was utilized to adjust hormone levels for day of cycle in an analysis of the association between total estradiol and risk of pre-menopausal breast cancer. A bootstrap method for obtaining variances of the logistic regression parameters which takes into account the error associated with estimation of the adjusted hormone levels has also been devised and will be implemented in the final year of the study. In addition, a method for analyzing correlated panel data using a continuous time Markov model is proposed. Correlated panel data arise when the states of each of several correlated clinical variables are repeatedly assessed over time for each individual. The proposed method can be applied to data from a clinical trial to simultaneously assess whether an experimental treatment for breast cancer has a greater toxic effect on several clinical variables compared with a standard treatment.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Biostatistics, Hormones, Repeated Measures, Measurement Error, Time-Dependent Covariates, Breast Cancer	51
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

NA In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

NA For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

NA In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

NA In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

NA In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature

2/25/98
Date

TABLE OF CONTENTS

Overview

This progress report describes research accomplished during the period December 31, 1996 - January 30, 1998. The manuscript describing the method for correcting for measurement error when subjects have a variable number of repeated measurements of exposure was published in **The American Journal of Epidemiology** (Kim, M.Y. and Zeleniuch- Jacquotte, A. "Correcting for measurement error in the analysis of case-control data with repeated measurements of exposure", American Journal of Epidemiology (1997) 145:1003-1010.) In addition, a paper describing the results of the application of the technique to a study of repeated measurements of androgens and risk of breast cancer in postmenopausal women appeared in the same journal. (Zeleniuch-Jacquotte, A., Bruning, P., Bonfrer, J., Koenig, K., Shore, R., Kim, M., Pastenack, B., Toniolo, P. "Relation of serum levels of testosterone and dehydroepiandrosterone sulfate with risk of breast cancer in postmenopausal women," American Journal of Epidemiology (1997) 145:1030-1038.) Reprints of the two articles are provided in the Appendix.

A method for analyzing correlated panel data was developed and a manuscript based on the results has been accepted by **Biometrics**. Correlated panel data can arise in studies where two or more correlated multi-state processes are periodically observed on each individual and the exact transition times between states are unknown. The data for a particular patient consist of repeated assessments over time of the states of each of several possibly correlated clinical variables. A procedure which models each process marginally under a time-homogeneous Markov model allowing for covariates was proposed. Although this technique was applied to data from an AIDS study, the method is also applicable to the analysis of toxicity data from clinical trials involving the treatment of breast cancer patients. Specifically, one can utilize the technique to simultaneously assess whether an experimental treatment for breast cancer has a greater toxic effect on several clinical variables compared with a standard treatment.

We also continued work on the method for adjusting for the systematic variability of hormone levels over the menstural cycle, based on a mixed ANOVA model with cubic splines. The method standardizes hormone measurements obtained at different time during the menstrual cycle, thus allowing for more valid comparisons of hormone levels between premenopausal breast cancer cases and controls. We have applied the method to adjust the hormone levels of premenopausal subjects in a nested case-control study of total estradiol and breast cancer risk. However, because of coding problems with the day-of-cycle data which subsequently became apparent, the methods will be re-applied to a revised data set. A bootstrap method for obtaining variances of the logistic regression parameters which takes into account the error associated with estimation of the calibration curve has been developed but not yet implemented.

Chapter I

Adjusting Hormone Levels for Day of Menstrual Cycle in Studies of Breast Cancer and Hormones in Pre-Menopausal Women

1 Introduction

Although levels of prolactin and bioavailable estradiol appear to be relatively stable over the phases of a woman's menstrual cycle, other hormones, such as total estradiol, fluctuate considerably. (Toniolo et al., 1993; Koenig et al., 1993.; Wu et al., 1976; Takatani et al., 1991). Epidemiologic studies investigating the association of total estradiol and risk of breast cancer among premenopausal women must adjust a subject's hormone level for day of cycle either in the design or analysis stage of the study, in order for the comparisons between cases and controls to be valid.

In the NYU Women's Health Study, a nested case-control study of serum hormonal levels and breast cancer, one of the criteria for matching controls with a breast cancer case among pre-menopausal women was the day of menstrual cycle on which the first blood specimen was collected, measured in number of days prior to next expected onset of menses. Subsequent blood donations, however, could not be matched on day of cycle. Therefore, a method was needed to standardize hormone measurements obtained at different times during the menstrual cycle for subjects in the same matched set.

Rosenberg et al (1994) used the first measurement from each control subject to fit a three-piece spline model to describe the change in total estradiol level over the menstrual cycle. For each subject, the estradiol measurement adjusted for day of cycle was then calculated as the difference between the observed value and the expected value from this calibration curve, measured in units of standard deviation. The limitation with this approach, however, is that because only the first measurement from each subject was used to fit the calibration curve, the curve is estimated with less precision than one that is estimated using all available repeated measurements. In addition, the width of the confidence intervals for the relative risks for breast cancer based on the adjusted estradiol measurements are underestimated, since they do not take into account the extra variation due to estimation of the parameters

of the calibration curve.

We propose an alternative method for describing the within-subject change in estradiol levels over the menstrual cycle, based on a mixed linear model with cubic splines, which utilizes all the repeated measurement data for each subject. The use of cubic splines in the model yields a smoother curve than the one fit by Rosenberg et al, which was based on a three-piece spline: two parabolas and a straight line, without smoothed join points.

We use the results from the mixed linear model to adjust each subject's hormone level for day-of-cycle. The adjusted measurement then becomes the exposure in a conditional logistic regression analysis. Bootstrap methods are utilized to obtain estimates of the corresponding 95% confidence intervals for the regression coefficients which account for the variation in the estimated calibration curve.

2 Methods

Let $\mathbf{y}_i = \{y_{i1}, ..., y_{ik_i}\}$ denote the vector of hormone levels for the i^{th} woman measured on k_i occasions for $i = 1, ...n$. Furthermore, let $\mathbf{t}_i = \{t_{i1}, ..., t_{ik_i}\}$ denote the vector of the number of days prior to next menses at which the $\mathbf{y}_i$ were measured. We assume a mixed linear model of the form

$$y_{ij} = \mu + \alpha_i + S(t_{ij}) + \epsilon_{ij}$$

where μ denotes an overall mean, α_i denotes a random subject effect from a $N(0, \sigma_\alpha^2)$ distribution, $S(t_{ij})$ is a cubic spline function, and the ϵ_{ij} are independent errors from a $N(0, \sigma_e^2)$ distribution. We further assume that the subject effects and the error terms are mutually independent. The above model implies that the correlation between repeated measurements of hormones on the same subject is equal to $\sigma_s^2/(\sigma_s^2 + \sigma_e^2)$.

We chose to use cubic splines to model estradiol levels versus day of cycle because this method provides great flexibility in fitting models, is visually smoth, and requires fewer constants to fit than higher degree splines. Rosenberg et al utilized two parabolic and one

linear function to describe the change in estradiol over the menstrual cycle, with only a single continuity restriction. Thus, although their overall function was continuous, it was not smooth at the two join points.

When fitting a cubic spline model, more join points or knots are better if the variable changes quickly over the covariate space. However, too many knots can lead to over-fitting of the data and more parameters to estimate. Stone (1986) suggested that 5 knots should provide enough flexibility for a reasonable number of degrees of freedom.

Given that the average length of a menstrual cycle is 28 days, we positioned 5 knots at the 5 day intervals: 5, 10, 15, 20, and 25 days prior to next menses. Using the + notation of Smith (1979), let

$$u_+ = u \quad \text{if} \quad u > 0$$

$$u_+ = 0 \quad \text{if} \quad u \leq 0.$$

Then the cubic spline can be specified as:

$$\begin{aligned} S(t) &= \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t-5)_+^3 + \beta_5 (t-10)_+^3 \\ &\quad + \beta_6 (t-15)_+^3 + \beta_7 (t-20)_+^3 + \beta_8 (t-25)_+^3. \end{aligned}$$

It follows that the overall mixed linear model has the following form:

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + \beta_4 (t_{ij}-5)_+^3 + \beta_5 (t_{ij}-10)_+^3 \\ &\quad + \beta_6 (t_{ij}-15)_+^3 + \beta_7 (t_{ij}-20)_+^3 + \beta_8 (t_{ij}-25)_+^3 + \epsilon_{ij} \qquad (1). \end{aligned}$$

This model assumes that the shape of the function describing the change in estradiol over the menstrual cycle is the same for all subjects, but that subjects can differ with regard to their baseline level on day 0.

Several techniques can be utilized to obtain estimates for the variance components and regression parameters in a mixed linear model, including traditional analysis of variance (ANOVA) methods, maximum likelihood (ML) methods, and restricted maximum likelihood

(REML) methods. REML estimates of the variance components are generally preferred, since ML estimates do not take into account the degrees of freedom used to estimate the fixed effects, which can result in estimates of variance components which are downwardly biased (Laird and Ware, 1982). The ANOVA methodology also has limitations, especially with unbalanced data, such as negative variance estimates, and lack of distributional properties (Searle et al). For these reasons, we used the REML method from the SAS PROC MIXED procedure to fit model (1).

Once the parameters in model (1) are estimated, estradiol levels adjusted for day of cycle can be computed using several approaches. One approach is to calculate the deviation of the subject's observed value from the expected value for that day of the cycle based on the fitted curve:

$$x_{ij} = y_{ij} - \hat{S}(t_{ij}). \tag{2}$$

Similarly, when repeated hormone measurements are available on all subjects and the average hormonal level is used as the exposure, the average adjusted for day of cycle can be calculated as:

$$\bar{x}_{i.} = \{\sum_j y_{ij} - \hat{S}(t_{ij})\}/n. \tag{3}$$

An alternative approach is to use the estimate of the random subject effect, $\hat{\alpha}_i$, from (1) for each $i = 1, ..., N$. The best linear unbiased predictor (BLUP) of α_i is $E(\alpha_i|\mathbf{y}_i, \hat{\beta}, \hat{\sigma}_s^2, \hat{\sigma}_e^2)$, the expected value of α_i, conditional on $\mathbf{y}_i, \hat{\beta}, \hat{\sigma}_s^2$, and $\hat{\sigma}_e^2$, which is also the empirical Bayes estimator of α_i. It can be shown that

$$E(\alpha_i|\mathbf{y}_i, \hat{\beta}, \hat{\sigma}_s^2, \hat{\sigma}_e^2) = \hat{R}_{n_i}(\bar{y}_i - \hat{S}(t_{ij})/n_i), \tag{4}$$

where

$$\hat{R}_{n_i} = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_s^2 + \frac{\hat{\sigma}_e^2}{n_i}}.$$

Note that R_{n_i} can also be interpreted as the reliability coefficient of $\bar{y}_i$.

Estradiol levels adjusted for day of cycle can be computed using one of the above approaches for all subjects and then used as the exposure in the usual logistic (or conditional logistic for matched studies) regression model to evaluate the association between estradiol level and risk of breast cancer among pre-menopausal women. The estimates of the standard errors for the regression parameter will be underestimated, however, since the uncertainty associated with the estimates for the calibration curve are not taken into account. In this paper, we utilize a bootstrap algorithm for obtaining estimates of the confidence intervals which include the variability contributed by estimation of the calibration curve. The algorithm is described in the next section.

3 Results

The primary aim of the NYU Women's Health Study is to determine whether endogenous hormones such as estradiol, are associated with risk of breast cancer. Between March 1985 and June 1991, a cohort of healthy women aged 34-65 years were enrolled at the Guttman Breast Diagnostic Institute, New York. At the time of enrollment and at annual screening visits thereafter, women were asked to donate blood and complete a self-administered questionnaire. Serum samples were frozen and stored for future biological assays. Subsequent cases of breast cancer were identified primarily through active follow-up and confirmed by reviewing medical and pathological records.

In order to limit the costs associated with measuring hormone levels in the cohort, a nested case-control study design was used. For each incident case of breast cancer, individually matched controls were selected at random from the risk set consisting of all cohort members alive and free of breast cancer at the time of diagnosis of the case, and who matched the case on menopausal status at entry, age at entry, and number and approximate dates of blood donations up to the date of diagnosis in the case. For additional details of the study design, see Toniolo et al (1991). The association between endogenous estrogens and breast

cancer in post-menopausal women was reported in Toniolo et al (1995). In this paper, we focus on the association between total estradiol and risk of breast cancer in pre-menopausal women.

A total of 498 estradiol measurements from 367 pre-menopausal control subjects were utilized to fit the calibration curve: 278 subjects had 1 measurement, 60 had 2, 28 had 3, and 4 had 4 measurements. Only measurements obtained less than 35 days prior to next menses were included. Total estradiol levels were log transformed prior to model fitting to adjust for deviations from normality.

The estimated mean curve describing the change in log estradiol level over the menstrual cycle is shown in Figure 1. The parameters in model (1) were estimated as follows: $\hat{\mu} = 4.16, \hat{\beta}_1 = .36, \hat{\beta}_2 = -.029, \hat{\beta}_3 = -.00067, \hat{\beta}_4 = .00061, \hat{\beta}_5 = -.012, \hat{\beta}_6 = .012, \hat{\beta}_7 = -.0077, \hat{\beta}_8 = .0052$. These estimates were then used to calculate levels of total estradiol adjusted for phase of menstrual cycle using the three approaches described above.

The results from fitting conditional logistic regression models to the adjusted total estradiol levels are shown in Table 1. The estimate of the logistic regression coefficient based on the adjusted first and average measurements calculated from (2) and (3), respectively, are similar to the estimate using the unadjusted first measurement because in the original study design, cases were matched to controls according to the phase of menstrual cycle of the first measurement.

In contrast, when the emprical Bayes estimator from (4) was utilized as the exposure, the regression coefficient estimate increased substantially. This increase is not surprising, given that the estimator in (4) can be viewed as an estimator of (3) that has been corrected for measurement error. Whittemore (1989), and Armstrong, Whittemore and Howe (1989), have proposed analogous forms of (4) as a method for correcting for measurement error in linear and logistic regression models. The method, commonly referred to as "Stein shrinkage", involves multiplying an exposure variable measured with error by the reliability of the exposure prior to fitting the regression model to obtain corrected coefficient estimates. In

the absence of confounders, measurement error in the exposure variable will result in relative risk estimates that are attenuated compared to the true relative risk. It follows that methods which correct for measurement error should yield higher estimates of relative risk than the uncorrected estimates, and that the regression coefficent based on the empirical Bayes estimator should be higher than the uncorrected average. It should be noted, however, that this estimator does not completely correct for measurement error because the reliability coefficient in (5) should be adjusted for the matching strata (ref. Kim et al, 1995). That is, the between-subject variance component may be overestimated if it is not computed controlling for the variation due to matching.

The width of the confidence intervals in Table (1) are underestimated because the extra variability due to estimation of the calibration curve is not taken into account. To obtain standard errors which incorporate this additional source of variation, we propose the following bootstrap procedure:

1. Generate a bootstrap sample from the control subjects.

2. Fit model (1) to the bootstrap sample to estimate the parmeters of the calibration curve

3. Generate a bootstrap sample from the matched cases and controls, using the matching stratum as the sampling unit.

4. Adjust the total estradiol measurements for day of menstrual cycle using the estimates from step (2).

5. Fit conditional logisitc regression models to the adjusted total estradiol measurements.

6. Repeat (1)-(4) 1,000 times, which is the approximate minimum number of bootstraps necessary to compute bias-corrected confidence limits.

The simple $(1-\alpha)\%$ confidence interval can be constructed using the $\alpha/2$ and $(1-\alpha/2)$ percentiles of the bootstrap distribution. Bias- corrected confidence intervals should be used when the bootstrap distribution of the regression parameter is asymmetric and when the sample size is small.

4 Conclusions

The computer program for generating the bootstrap sample described above has been developed using the SAS macro facility, but still needs to be tested. In addition, progress on this project has been slower than anticipated because of problems with the coding of the day-of-cycle data which needed to be resolved. For example, Figure 1 shows that an unusually large number of subjects gave a blood sample 20 days prior to the next menstrual cycle. This is an artifact of the coding rule that cohort members with irregular or unknown cycle lengths and whose blood donation occurred between day 1 and day 8 after the first day of last menustrual cycle were considered to have been in the early follicular phase and at 20 days prior to next menses. Because this coding mechanism could adversely impact the fit of the model in the area of the curve to the right of 20 days prior to next menses in Figure 1, model (1) needs to be refitted to the data after identifying and excluding the measurements which were artificially assigned to the time of 20 days prior to next menses. We anticipate that this will be accomplished by the end of this final year of the funding period, along with implementation of the bootstrap method and preparation of a manuscript describing the methods and results.

References

Amstrong, B., Whittemore, A., and Howe, G. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. Statistics in Medicine (1989) 8:1151-1163.

Koenig, K.L., Toniolo, P.G., Bonfrer, P.F., et al. Reliability of serum prolactin measurements. Cancer Epidemiology, Biomarkers and Prevention (1993) 137:1068-1080.

Rosenberg, C.R., Pasternack, B.S., Shore, R.E., et al. Premenopausal estradiol levels and the risk of breast cancer: a new method of controlling for day of the menstrual cycle. Am J Epidemiol (1994) 140: 518-525.

Smith, P.L. Splines as a useful and convenient tool. The American Statistician (1979) 33: 57-62.

Stone, C.J. Comment on Hastie and Tibshirani. Statistical Science (1986) 1: 312-34.

Takatani, O., Okumoto, T., Kosano, H. Genesis of breast cancer in Japanese: A possible relationship between sex hormone binding globulin (SHBG) and serum lipid components. Breast Cancer Res Treat (1991) 18:s27-s29.

Toniolo, P., Pasternack, B., Shore, R., Sonnenschein, e., Koenig, K., Rosenberg, C., Strax, P., Strax, S. Endogendous hormones and breast cancer: A prospective cohort study. Breast Cancer Res Treat (1991) 18:S23-S26.

Toniolo, P., Levitz, M., Zeleniuch-Jacquotte, A., Banerjee, S., Koenig, K., Shore, R., Strax,

P., Pasternack, B. A prospective study of endogenous estrogens and breast cancer in post-menopausal women. Journal of the National Cancer Institute (1995) 87-190-197.

Wu, C.H. Free and protein-bound plasma estradiol-17B during the menstrual cycle. J Clin Endocrinol Metab (1976) 43:436-445.

Whittemore, A.S. Errors-in-variables regression using Stein estimates. The American Statistician (1989); 43:226-228.

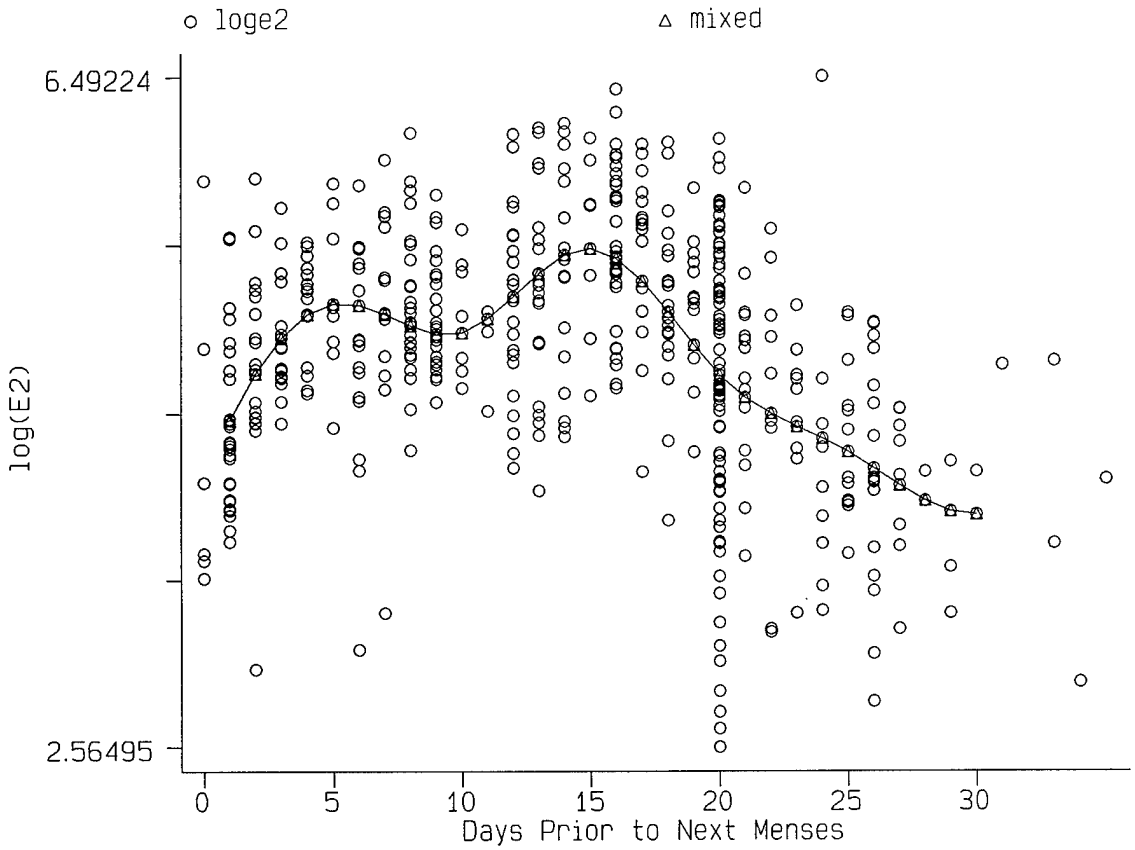
o loge2
Δ mixed
6.49224
2.56495
log(E2)
0
5
10
15
20
25
30
Days Prior to Next Menses

Table 1: Logistic Regression Parameter Estimates and 95% Confidence Intervals for the Associations of Total Estradiol Level and Risk of Breast Cancer in Pre-menopausal Women

Exposure Variable	Regression Coefficient	95%C.I.
Total Estradiol		
First measurement	0.19	(-0.23 - 0.61)
Adjusted first measurement	0.26	(-0.17 - 0.70)
Adjusted average	0.17	(-0.28 - 0.63)
Empirical Bayes adjusted average	1.52	(-0.83 - 3.87)

Chapter II

The Analysis of Correlated Panel Data Using a Continuous Time Markov Model

1 Introduction

Many biological processes may be described in terms of a finite number of states which individuals proceed through over time. For example, Klein et al (1984) and Kay (1986) have modeled the phases of cancer as a three stage disease process.

In most longitudinal studies which follow the passage of individuals through various biological states, the available data consist of the series of states each subject was observed to be in at various points in time. Because monitoring is not continuous, information about the process is usually unavailable between follow-up times, and the exact times of transition from one state to another are not known. This type of data, in which the observations consist of the states occupied by the individuals under study at a sequence of discrete time points, is often referred to as "panel" data. Methods for analyzing panel data using time-homogeneous Markov models have been explored by Klein et al. (1984), Kalbfleisch and Lawless (1985), and Gentleman et al (1994).

Existing methods for the analysis of panel data have been applicable to the case where a single biological process is of interest. However, many clinical studies monitor two or more processes over time on each individual, where the processes observed within a subject may be correlated. For example, in certain longitudinal studies involving HIV-infected subjects, several immunologic variables, such as CD4, CD8 and serum immunoglobulins are each measured periodically on all subjects. Also, in most clinical trials, several possibly correlated clinical variables are repeatedly assessed during the follow-up period of each patient.

In the clinical trial which motivated the methodology proposed in this paper, patients with the Acquired Immune Deficiency Syndrome (AIDS) were randomized to receive one of two drugs: ganciclovir or foscarnet, for the treatment of cytomegalovirus (CMV) retinitis, and were periodically monitored for the development of toxicities in several hematologic, enzymatic, and chemical variables. For each variable, five states of toxicity were defined: none, mild, moderate, severe and life-threatening, according to the AIDS Clinical Trials Group (ACTG) toxicity grading system. The data for a

particular patient consisted of repeated assessments over time of the toxicity states of each of several possibly correlated clinical variables, yielding what we refer to as "correlated panel" data. To our knowledge, statistical methods for analyzing correlated panel data have not been previously developed.

Wei, Lin and Weissfeld (1989) described a method for analyzing multivariate survival data which models each survival distribution marginally using a Cox proportional hazards model, and makes no specific assumptions regarding the structure of dependence between distinct failure times on each subject. This approach has also been utilized by Stram, Wei and Ware (1988). In this paper, we apply the marginal approach to the analysis of correlated panel data. We model each process based on a time-homogeneous Markov model allowing for covariates and impose no specific dependence structure among the related processes. The resulting estimators are shown to be asymptotically jointly normal with a covariance matrix that accounts for the dependence among related processes and can be consistently estimated. Simultaneous inference procedures are also proposed. In Section 2, the method for analyzing correlated panel is described, and in Section 3 an example is presented using data from the CMV clinical trial.

2 Methods

2.1 Markov Modeling and Estimation

Assume that K processes are periodically observed on each subject, and that each process has Q ordered states. We model the k^{th} process, $(k = 1, ..., K)$, as a time-homogeneous Markov process, where the first $Q-1$ are transient states, and the Qth is a single absorbing state. Transitions are assumed to occur irreversibly from one state to the next. The time-homogeneous Markov assumption implies that times to transition between states are exponentially distributed with a hazard, or transition intensity, that may be modeled as a function of covariates.

For notational convenience, we assume a single covariate, z, is of interest. Then, following the proportional hazards model of Cox (1972), the transition intensities for each process can be modeled as,

$$\lambda_q^k(z) = \lambda_q^k \exp(\beta_q^k z), \qquad q = 1, ..., Q-1, k = 1, ..., K$$

where λ_q^k is the baseline transition intensity between states q and $q+1$ for the kth process, and β_q^k is the regression parameter denoting the effect of z on λ_q^k.

For the i^{th} subject ($i = 1, ..., n$) with covariate z_i, the probability that the k^{th} process will be in state s at time t_1+t, given that the process was in state r at time t_1, is denoted as $p_{rs}^k(t, z_i)$. Then, under the assumptions that $\lambda_j^k(z_i) \neq \lambda_l^k(z_i)$, $l \neq j, j, l = 1, ..., Q-1$, standard methods may be utilized to show that this transition probability is equal to

$$p_{rs}^k(t, z_i) = (-1)^{s-r} \lambda_r^k(z_i) ... \lambda_{s-1}^k(z_i) \sum_{j=r}^{s} \left[\frac{e^{-\lambda_j^k(z_i)t}}{\prod_{l=r, l\neq j}^{s} \{\lambda_j^k(z_i) - \lambda_l^k(z_i)\}} \right], \qquad (1)$$

where $r = 1, ..., Q-1; s \geq r$, and $s < Q$ (Chiang, 1980). The transition probabilities from a transient state r to the absorbing state Q are equal to

$$p_{rQ}^k(t, z_i) = 1 - \sum_{j=r}^{Q-1} p_{rj}^k(t, z_i), \; r = 1, ...Q-1,$$

and $p_{QQ}^k(t, z_i) = 1$.

Now, let s_{ij}^k denote the state of the kth process for the ith subject at the jth observation ($j = 1, ., ., m_i$), at time t_{ij}^k, measured in days since date of initial observation. Then the observed history of the k^{th} process for the i^{th} subject may be denoted as

$$A_i^k = (s_{i1}^k, t_{i1}^k, ..., s_{im_i}^k, t_{im_i}^k).$$

Following the Markov property and conditional probability arguments, the probability of observing A_i^k, conditional on z_i, can be expressed as,

$$\begin{aligned} \Pr(A_i^k | z_i) &= \Pr(k^{th} \text{ process in state } s_{i1}^k \text{ at time } t_{i1}^k, ..., \text{state } s_{im_i}^k \text{ at time } t_{im_i}^k | z_i) \\ &= \prod_{j=1}^{m_i - 1} p_{s_{ij}, s_{i,j+1}}^k (t_{i,j+1}^k - t_{ij}^k, z_i), \end{aligned}$$

where the form of $p^k_{s_{ij},s_{i,j+1}}(t^k_{i,j+1}-t^k_{ij},z_i)$ may be derived from (1). Then the marginal likelihood function for the k^{th} process over all n subjects is

$$\begin{aligned} L(\theta_k) &= \prod_{i=1}^{n} Pr(A_i^k|z_i) \\ &= \prod_{i=1}^{n}\left\{\prod_{j=1}^{m_i-1} p^k_{s_{ij},s_{i,j+1}}(t^k_{i,j+1}-t^k_{ij},z_i)\right\}, \end{aligned}$$

where $\theta_k=(\lambda_1^k,...\lambda_{Q-1}^k\beta_1^k,...\beta_{Q-1}^k)$. The corresponding log-likelihood is equal to

$$\mathcal{L}(\theta_k)=\sum_{i=1}^{n}\left[\sum_{j=1}^{m_i-1}\log\{p^k_{s_{ij},s_{ij+1}}(t^k_{ij+1}-t^k_{ij},z_i)\}\right]. \tag{2}$$

Substituting (1) in (2), the maximum likelihood estimator, $\hat{\theta}_k$, of θ_k, for each process may be obtained by solving $\partial\mathcal{L}(\theta_k)/\partial\theta_k=0$, for $k=1,...,K$. It follows that under the usual regularity conditions, $\sqrt{n}(\hat{\theta}_k-\theta_k)$ for each process is asymptotically normal with mean 0 and covariance matrix $I_k^{-1}(\theta_k)$, where $I_k(\theta_k)$ is defined as

$$I_k(\theta_k)=E_{\theta_k}\left[-\frac{1}{n}\frac{\partial^2\mathcal{L}(\theta_k)}{\partial\theta_k^2}\right]=E_{\theta_k}\left[\frac{1}{n}\left\{\frac{\partial\mathcal{L}(\theta_k)}{\partial\theta_k}\right\}\left\{\frac{\partial\mathcal{L}(\theta_k)}{\partial\theta_k}\right\}^T\right]. \tag{3}$$

The estimates of θ_k from different processes may be correlated. To take into account the correlations among $\hat{\theta}_k$'s, consider the Taylor series expansion

$$\sqrt{n}(\hat{\theta}_k-\theta_k)\sim\mathcal{I}_k^{-1}(\theta_k)\frac{U_k}{\sqrt{n}},$$

where

$$\frac{1}{\sqrt{n}}U_k=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial\mathcal{L}_i(\theta_k)}{\partial\theta_k}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_{ik}, \tag{4}$$

and

$$\mathcal{I}_k(\theta_k)=-\frac{1}{n}\frac{\partial^2\mathcal{L}(\theta_k)}{\partial\theta_k^2}.$$

By the above development, $\sqrt{n}\left\{\begin{pmatrix}\hat{\theta}_1\\ \vdots\\ \hat{\theta}_K\end{pmatrix}-\begin{pmatrix}\theta_1\\ \vdots\\ \theta_K\end{pmatrix}\right\}$ can be written as the sum of independent vectors with mean 0, and thus, under the usual regularity conditions,

follow a multivariate normal distribution with mean 0 and covariance matrix Σ, which may be partitioned as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1K} \\ \Sigma_{12} & \Sigma_{22} & \cdots & \Sigma_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1K} & \Sigma_{2K} & \cdots & \Sigma_{KK} \end{bmatrix},$$

where $\Sigma_{kk} = I_k^{-1}(\theta_k)$, and Σ_{jk} is the covariance matrix between $\sqrt{n}(\hat{\theta}_j - \theta_j)$ and $\sqrt{n}(\hat{\theta}_k - \theta_k)$. From (4), the covariance matrix of $(U_k/\sqrt{n},\ U_j/\sqrt{n})$ can be estimated consistently by

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\xi}_{ik}\hat{\xi}_{ij}^T, \tag{5}$$

where $\hat{\xi}_{ik}$ is equal to $\partial \mathcal{L}_i(\theta_k)/\partial\theta_k$ evaluated at $\hat{\theta}_k$, and $\mathcal{L}_i(\theta_k)$ is the contribution to the log-likelihood made by the i^{th} subject. The term $\hat{\xi}_{ij}$ is defined similarly. Since $\mathcal{I}_k^{-1}(\hat{\theta}_k)$ and $\hat{I}_k^{-1}(\hat{\theta}_k) = (\frac{1}{n}\sum_{i=1}^{n} \hat{\xi}_{ik}\hat{\xi}_{ik}^T)^{-1}$ are equivalent in probability, it follows that one can estimate Σ_{jk} by

$$\frac{1}{n}\hat{I}_j^{-1}(\theta_j)\left(\sum_{i=1}^{n} \hat{\xi}_{ij}\hat{\xi}_{ik}^T\right)\hat{I}_k^{-1}(\theta_k),$$

for $j \neq k$ and $j, k = 1, ..., K$.

2.2 Hypothesis Testing

The multivariate normal distribution of $(\hat{\theta}_1, ..., \hat{\theta}_K)$ provides a basis for making simultaneous inferences about the θ_ks. In particular, suppose that we are interested in evaluating the effect of z_i on the transition intensities for the K processes. Let $\eta = \{\eta_1, ..., \eta_{K\times(Q-1)}\} = (\beta_{11}, ..., \beta_{Q-1,1}, ..., \beta_{1K}, ..., \beta_{Q-1,K})$ denote the parameters of interest, and let $\hat{\Phi}$ denote the estimate of the covariance matrix of $\hat{\eta}$, which can be obtained from $\hat{\Sigma}$. Then asymptotically the following holds

$$W = \hat{\eta}'\hat{\Phi}^{-1}\hat{\eta} \sim \chi^2_{\{K\times(Q-1)\}}. \tag{6}$$

The above W statistic can then be used to simultaneously test the null hypotheses H_l: $\eta_l = 0$, for $l = 1, ..., K \times (Q-1)$. If the null hypotheses are rejected, the next step is to decide which of the η_l are non-zero.

Let $\{\tilde{\eta}_1, \tilde{\eta}_2, ..., \tilde{\eta}_{K\times(Q-1)}\}$ denote the standardized estimator of $\{\hat{\eta}_1, \hat{\eta}_2, ..., \hat{\eta}_{K\times(Q-1)}\}$, where $\tilde{\eta}_l = \hat{\eta}_l/\hat{\phi}_{ll}^{1/2}$, and $\hat{\phi}_{ll}$ is the $(l,l)^{th}$ element of $\hat{\Phi}$. Under the null hypotheses H_l, $\{\tilde{\eta}_1, ..., \tilde{\eta}_{K\times(Q-1)}\}$ is approximately normal with mean 0 and covariance matrix $\tilde{\Phi} = \{\tilde{\phi}_{pq}\}$, where $\tilde{\phi}_{pq} = \hat{\phi}_{pq}/(\hat{\phi}_{pp}\hat{\phi}_{qq})^{1/2}$, and $\hat{\phi}_{pq}$ is the (p,q)th element in $\hat{\Phi}$. A conventional multiple testing procedure rejects $H_l, l = 1, ..., K \times Q - 1$, if $|\tilde{\eta}_l| > c$, where c is the smallest number such that

$$Pr(|\tilde{\eta}_l| < c, l = 1, ..., K \times Q | H_1, ..., H_{K\times Q-1}) \leq 1 - \alpha,$$

and α is a prespecified level of signifcance. However, the sequential multiple test procedures studied by Marcus, Peritz and Gabriel (1976), Holm (1979) and Wei and Stram (1988) will yield more powerful tests than the above conventional multiple testing procedures.

Following Wei and Stram (1988), let $\tilde{\eta}_l^*$ be the lth largest absolute value of the $\tilde{\eta}_l$'s, and let $\tilde{\Phi}^*$ be the corresponding variance-covariance matrix that is obtained by rearranging the rows and columns of $\tilde{\Phi}$. Also, let $H_l^* : \eta_l^* = 0$ be the ordered hypotheses from the H_l's according to the order of $\tilde{\eta}_1^*, ..., \tilde{\eta}_{K\times Q-1}^*$. Furthermore, let $(Z_1, ..., Z_{K\times Q-1})$ denote a multivariate normal vector with mean zero and covariance matrix $\tilde{\Phi}^*$. Starting with the hypothesis H_1^*, we reject $H_l^*, l = 1, ..., K \times Q - 1$, if $P(\min_{l\leq j\leq K\times Q-1} Z_j \leq -|\tilde{\eta}_l^*|) \leq \alpha/2$, where α is a prespecified two-sided level of significance, provided that $H_1^*, ..., H_{l-1}^*$ have been tested and rejected. It can be shown that the asymptotic type I error probability of this procedure is α for any combination of true H_ls. An illustration of the sequential multiple testing procedure is provided in the example.

3 Example

The methods developed in the previous section were applied to data derived from the CMV retinitis clinical trial. This trial, conducted by the Studies of Ocular Complications of AIDS Research Group, was designed to compare ganciclovir (Cytovene, Syntex Laboratories) and foscarnet (Foscavir, Astra Pharmaceutical Products) in the treatment of CMV retinitis in AIDS patients (SOCA, 1992).

Two hundred and thirty-four patients were randomly assigned to treatment: 127 to ganciclovir and 107 to foscarnet. One of the objectives of this trial was to compare the toxicity profiles of the two drugs. We focus particularly on the toxic effects of the treatments on the two hematologic variables: hemoglobin and absolute neutrophil count, which could be correlated within individuals. Five stages of toxicity were defined for each outcome based on the following ACTG toxicity criteria. For HGB, the stages were defined as, none: $\geq 11.0 g/dl$, mild: $9.5 - 10.9 g/dl$, moderate: $8.0 - 9.4 g/dl$, severe: $6.5 - 7.9 g/dl$ and life-threatening:$< 6.5 g/dl$. The corresponding toxicity grades for ANC were, none: $\geq 1500/\mu l$, mild: 1000-1499$/\mu l$, moderate: 750-999$/\mu l$, severe: $500 - 749/\mu l$, and life-threatening: $< 500/\mu l$. It should be emphasized that even though the quantitative levels of each variable were known, the grade of toxicity rather than the actual level was more clinically relevant. In addition, because very few subjects were observed to progress to the life-threatening stage of both HGB and ANC toxicity, this state was combined with the severe toxicity state to yield four possible states for each process. The goals of the analysis were to estimate and compare the effects of foscarnet and ganciclovir on the transition intensities between the states of toxicity of HGB and ANC, and to estimate the average waiting time in each state.

The methods in the previous section were based on the assumption that transitions between states occur irreversibly. Levels of HGB and ANC may not decrease monotonically over time, but could increase from one visit to the next. When this occurred in our data set, the process was simply held in the state defined by the previous observation until it was observed to decline again, thus allowing us to model the states as irreversible. The frequencies of occurrences in which the processes were observed to increase rather than decrease from one visit to the next were relatively low and were similar in the two groups (ANC: 20% vs. 18% for foscarnet and ganciclovir, respectively; HGB: 21% vs. 18% for foscarnet and ganciclovir, respectively), so that the assumption of the progressive model would not likely bias the estimates of treatment effects.

The covariate denoting treatment group for the i^{th} subject is defined as $z_i = 0$

for Foscarnet, and $z_i = 1$ for Ganciclovir. The parameter vector for HGB is denoted as $\theta_1 = (\lambda_1^1, \lambda_2^1, \lambda_3^1, \beta_1^1, \beta_2^1, \beta_3^1)$, and the corresponding vector for ANC is $\theta_2 = (\lambda_1^2, \lambda_2^2, \lambda_3^2, \beta_1^2, \beta_2^2, \beta_3^2)$, where β_q^k is the coefficient corresponding to the effect of treatment on the transition intensity between states q and $q+1$ for process k.

Tables 1 and 2 show the estimates of the transition intensities and the β coefficients for the two processes. The mean waiting time for the k^{th} state was estimated by $1/(\hat{\lambda}e^{z\hat{\beta}})$. The mean waiting times in each state of ANC for ganciclovir are less than half the mean waiting times for foscarnet, indicating the toxic effects on ANC are more severe with ganciclovir. On the other hand, for HGB, foscarnet appears to be associated with a higher probability of transition from mild to moderate, and moderate to severe toxicity.

Figures 1 and 2 show the estimated cumulative distribution functions of time to severe toxicity of HGB and ANC, respectively, for each treatment. The difference between treatments in the probability of reaching severe toxicity is much greater for ANC than HGB. The Kaplan-Meier estimates of the distribution functions are also shown for comparison. Our model appears to underestimate the probabilities of progressing to severe toxicity relative to the Kaplan-Meier estimates. The lack-of-fit of the model was also apparent in results (not shown) of the goodness-of-fit test proposed by Gentleman (1994) that was performed. The inadequacy of the assumed model could indicate that the estimates of the regression coefficients and corresponding covariance matrix may not be valid; however, for illustrative purposes we proceed to present the results of the sequential multiple testing procedure.

The covariance matrix for $(\hat{\beta}_1^1, \hat{\beta}_2^1, \hat{\beta}_3^1, \hat{\beta}_1^2, \hat{\beta}_2^2, \hat{\beta}_3^2)$ was estimated to be

$$\begin{pmatrix} 0.0560 & 0.0020 & 0.0023 & 0.0006 & 0.0116 & -0.0009 \\ & 0.0689 & -0.0006 & 0.0087 & 0.0072 & -0.0004 \\ & & 0.1072 & 0.0088 & -0.0023 & 0.0082 \\ & & & 0.0861 & 0.0024 & -0.0014 \\ & & & & 0.0692 & 0.0082 \\ & & & & & 0.0968 \end{pmatrix}.$$

To jointly test $H_l^k : \beta_l^k = 0, l = 1,2,3; k = 1,2$, we calculated the test statistic, W, from (4), which approximately follows a χ^2 distribution with 6 degrees of freedom. The observed W was 42.96 which is significant at less than 1% level.

In order to evaluate which of the coefficients were non-zero, we employed the multiple testing procedure described in the previous section. The standardized parameter estimates $(\tilde{\beta}_1^1, \tilde{\beta}_2^1, \tilde{\beta}_3^1, \tilde{\beta}_1^2, \tilde{\beta}_2^2, \tilde{\beta}_3^2)$ were 3.477, 3.910, 2.629, -1.721, -1.039, and 0.426 respectively. Since $-|\tilde{\beta}_2^1| < -|\tilde{\beta}_1^1| < -|\tilde{\beta}_3^1| < -|\tilde{\beta}_1^2| < -|\tilde{\beta}_2^2| < -|\tilde{\beta}_3^2|$, we tested the null hypotheses in the order: $H_2^1, H_1^1, H_3^1, H_1^2, H_2^2, H_3^2$. Using the integration algorithm for multivariate normal probabilities in Schervish(1984, 1985), we have

$$\begin{aligned} P(\min\{Z_1, Z_2, Z_3, Z_4, Z_5, Z_6\} \leq -3.910) &\approx 0.0002 \\ P(\min\{Z_2, Z_3, Z_4, Z_5, Z_6\} \leq -3.477) &\approx 0.0012 \\ P(\min\{Z_3, Z_4, Z_5, Z_6\} \leq -2.629) &\approx 0.0279 \\ P(\min\{Z_4, Z_5, Z_6\} \leq -1.721) &\approx 0.1230. \end{aligned}$$

Assuming a two-sided $\alpha = 0.05$, these results suggest that ganciclovir has a greater toxic effect on ANC than foscarnet, across all states of toxicity. However, the evidence is insufficient to conclude that the toxicity profiles for HGB are different for the two treatments.

4 Conclusions

Although the proposed methods were applied to data from a clinical trial of two treatments for CMV retinitis, they are also applicable to longitudinal studies of other diseases, such as breast cancer, in which several clinical variables are periodically monitored over time on each subject.

A key assumption of the methods is that transition rates between states do not change over time. This time-homogeneous Markov assumption is quite restrictive, however, and may in part explain the apparent lack of fit of the model. An alternative approach is to assume that the intensities are piecewise constant over distinct time

intervals (Gentleman et al., 1994). Kalbfleisch and Lawless (1985) also suggested fitting a parametric time-dependent model.

In addition, we have assumed a progressive Markov model, in which transitions between states can only occur in one direction. The proposed methodology can in principle be extended to the situation where arbitrary transitions are allowed, by modelling each process marginally using the methods in Kalbfleisch and Lawless (1985) and deriving the joint covariance matrix estimate of all model parameters empirically. However, because a larger number of parameters are required to specify the reversible model compared with the progressive model, estimation of the parameters may be more difficult.

ACKNOWLEDGEMENTS

This research was supported by Grants AI-27742 and DA-04722 from the National Institutes of Health, and DAMD17-94-J-4174 from the Department of Defense. We are grateful to the Studies of the Ocular Complications of AIDS Research Group (SOCA) for providing the data in the example. The analyses and interpretations herein are solely those of the authors and not of the SOCA research group as a whole, and representatives of the SOCA Research Group have not reviewed or checked the analyses or interpretations so presented for clarity or accuracy or for compliance with the stated principles of SOCA.

REFERENCES

Chiang, C.L. (1980) *An Introduction to Stochastic Processes and Their Applications,* 2nd edn, Krieger, New York.

Cox, D.R.(1972), "Regression Models and Life-Tables"(with discussion), *Journal of the Royal Statistical Society,* Ser. B. 34, 187-220.

Gentleman, R.C., Lawless, J.F., Lindsey, J.C. and Yan, P.(1994), "Multi-State Markov Models For Analysing Incomplete Disease History Data With Illustraions For HIV Disease," *Statistics in Medicine*, 13, 805-821.

Holm, S.(1979), "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6, 65-70.

Kalbfleisch, J.D. and Lawless, J.F.(1985), "The Analysis of Panel Data Under A Markov Assumption," *Journal of the American Statistical Association*, 80, 863-871.

Kay, R.(1986), "A Markov Model for Analysing Cancer Markers and Disease States in Survival Studies," *Biometrics*, 42, 855-865.

Klein, J.P., Klotz, J.H. and Grever, M.R.(1984), "A Biological Marker Model for Predicting Disease Transitions," *Biometrics*, 40, 927-936.

Longini Jr. I.M., Clark, W.S., Byers, R.H., Ward, J.W., Darrow, W.W., Lemp, G.F. and Hethcote, H.W.(1989), "Statistical Analysis of the Stages of HIV Infection Using a Markov Model," *Statistics in Medicine*, 8, 831-843.

Marcus, R., Peritz, E. and Gabriel, K.R.(1976), "On Closed Testing Procedures with Special Reference to Ordered Analysis of Variance," *Biometrika*, 63, 655-660.

Schervish, M.J. (1984), "Multivariate Normal Probabilities with Error Bound, "*Applied Statistics*, 33, 81-94.

Stram, D.O., Wei, L.J., and Ware, J.H.(1988), "Analysis of Repeated Ordered Categorical Outcomes with Possibly Missing Observations and Time-Dependent Covariates," *Journal of the American Statistical Association*, 83, 631-637.

Studies of Ocular Complications of AIDS Research Group (1992), "Mortality in Patients With the Acquired Immunodeficiency Syndrome Treated With Either Foscarnet or Ganciclovir For Cytomegalovirus Retinitis," *The New England Journal of Medicine*, 326:213-220.

Wei, L.J., Lin, D.Y. and Weissfeld, L.(1989), "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distribution," *Journal of the American Statistical Association*, 84, 1065-1073.

Wei, L.J. and Stram, D.O.(1988), "Analysing Repeated Measurements with Possibly Missing Observations by Modeling Marginal Distributions," *Statistics in Medicine*, 7, 139-148.

Table-1. Estimated parameters and mean waiting times in each state of toxicity for Absolute neutrophil count

Stage	($\lambda \pm$S.D.,$\beta \pm$S.D.)	Mean waiting time for Fos.(days)	Mean waiting time for Gov.(days)
1	(.0070±.0013, .823±.237)	(143 ± 27)	(63 ± 8)
2	(.0059±.0014, 1.03±.263)	(169 ± 40)	(61 ± 8)
3	(.0065±.0020, .861±.327)	(154 ± 47)	(65 ± 8)

Table-2. Estimated parameters and mean waiting times in each state of toxicity for Hemoglobin count

Stage	($\lambda \pm$S.D.,$\beta \pm$S.D.)	Mean waiting time for Fos.(days)	Mean waiting time for Gov.(days)
1	(.0139±.0027, -.505±.293)	(72 ± 14)	(119 ± 26)
2	(.0098±.0018, -.273±.263)	(102 ± 19)	(135 ± 25)
3	(.0062±.0013, 0.133±.311)	(161 ± 34)	(142 ± 32)

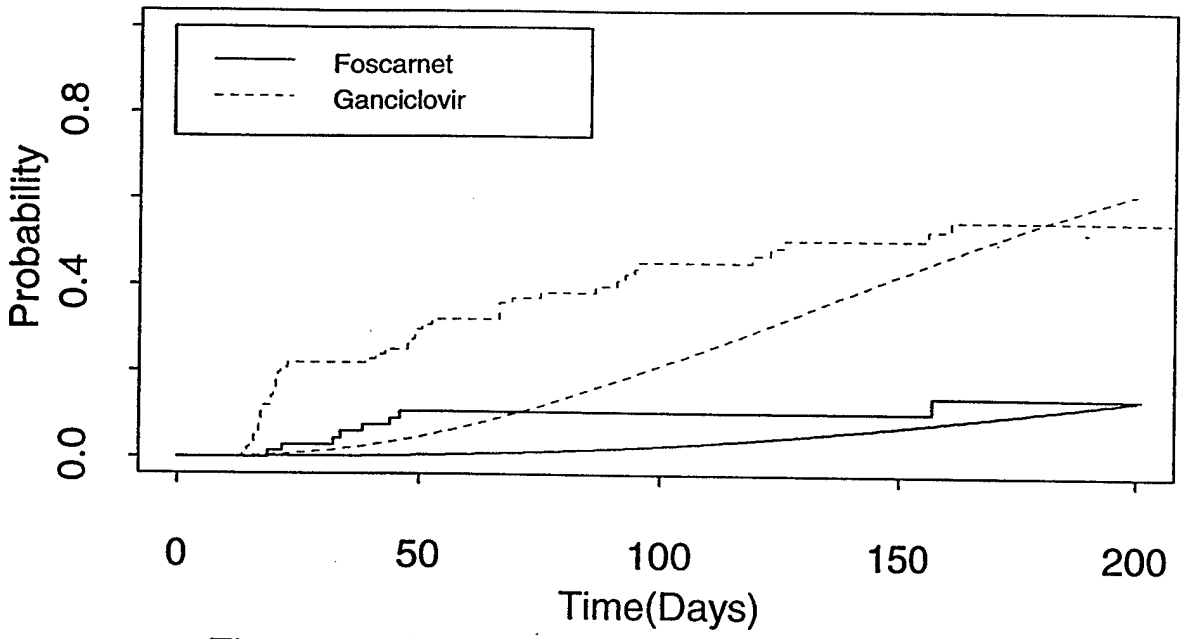

Figure 1. C.D.F. and Kaplan-Meier Estimates of Time to Severe Toxicity of ANC

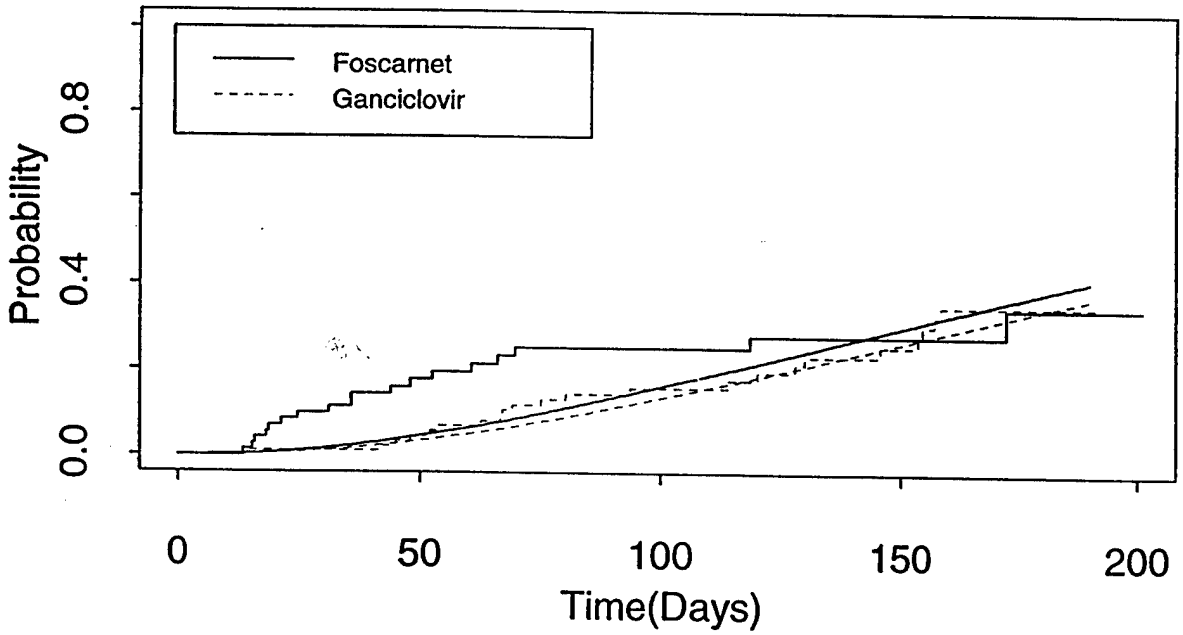

Figure 2. C.D.F. and Kaplan-Meier Estimates of Time to Severe Toxicity of HGB

Appendix

American Journal of Epidemiology

Vol. 145, No. 11
Printed in U.S.A.

Correcting for Measurement Error in the Analysis of Case-Control Data with Repeated Measurements of Exposure

Mimi Y. Kim and Anne Zeleniuch-Jacquotte

The authors present a technique for correcting for exposure measurement error in the analysis of case-control data when subjects have a variable number of repeated measurements, and the average is used as the subject's measure of exposure. The true exposure as well as the measurement error are assumed to be normally distributed. The method transforms each subject's observed average by a factor which is a function of the measurement error parameters, prior to fitting the logistic regression model. The resulting logistic regression coefficient estimate based on the transformed average is corrected for error. A bootstrap method for obtaining confidence intervals for the true regression coefficient, which takes into account the variability due to estimation of the measurement error parameters, is also described. The method is applied to data from a nested case-control study of hormones and breast cancer. *Am J Epidemiol* 1997;145:1003–10.

epidemiologic methods; breast neoplasms; hormones; measurement error; statistics

In most case-control studies, the risk factors of interest are measured with error. For biologic variables, such as blood pressure, nutrients, and hormone levels, measurement error can arise from limitations in the measurement technique or laboratory assay. In addition, because the exposure of interest is usually a subject's underlying long-term average value rather than the level at any single point in time, intrinsic fluctuations in the variable over time can also contribute to measurement error.

When the error is random and nondifferential with respect to case-control status, it is well known that estimates of relative risk based on the mis-measured exposure will be attenuated. In order to minimize the effects of measurement error, many investigators advocate collecting repeated measurements of the exposure on all subjects and using the individual's average value (1). However, as noted by Rosner et al. (2), even when the mean of several replicates is substituted for a single measurement, attenuation of relative risks may still occur, especially when the degree of measurement error is large and the average is based on only a few repeats.

Methods for correcting estimates of relative risk for measurement error have been proposed in a number of epidemiologic and statistical papers (3, 4). The most common method involves correcting the "naive" relative risk estimate based on the observed exposure by the expected amount of bias. In the case of logistic regression, the regression parameter will be attenuated by the factor, R, which is equal to the reliability coefficient of the mis-measured exposure (1, 2). Therefore, one can multiply the biased estimate of the regression coefficient by the inverse of the reliability coefficient to obtain the corrected estimate. This method, however, is dependent on the assumption that the reliability of the exposure measurement is the same for all subjects. When the average of several replicates is used as the measure of exposure, this condition will be met only if all subjects have an equal number of repeated measurements, given that the degree of measurement error associated with a single measurement is the same for all subjects.

In studies in which the exposure is measured on repeated occasions, however, subjects often have a variable number of measurements because of missing data. For example, the data that are utilized to illustrate the methods in this paper are derived from a nested case-control study of serum hormonal levels and breast cancer from the New York University (NYU) Women's Health Study (5). The study cohort consists of 14,275 women who donated multiple blood samples over time and have been followed since enrollment for the development of breast cancer. Most women have donated one or two samples; however, many have also donated three or more samples. Because subjects with

Received for publication March 18, 1996, and accepted for publication January 29, 1997.

Abbreviations: NYU, New York University; SHBG, sex hormone-binding globulin.

From the Epidemiology and Biostatistics Program, Institute of Environmental Medicine and Kaplan Comprehensive Cancer Center, NYU Medical Center, New York, NY.

a larger number of multiple blood samples have a more reliable estimate of their true underlying serum hormonal levels than subjects with fewer measurements, the reliability of the measured exposure will not be constant across individuals. Consequently, the usual procedure for correcting for measurement error cannot be applied.

Liu and Liang (6) proposed an estimating equation approach for obtaining consistent estimates of logistic regression parameters when all subjects have the same number of repeated imprecise exposure measurements, which in principle could be extended to the more complicated situation when the number of replicates is variable between subjects. In this paper, we discuss an alternative method for correcting for measurement error in the analysis of matched case-control data when subjects have a variable number of repeated exposure measurements and the individual's average is used as the measure of exposure. The technique, which assumes that both the true exposure and the measurement error are normally distributed, involves multiplying each subject's observed average by the reliability of the average prior to fitting the logistic regression model. The resulting logistic regression coefficient based on the transformed average is corrected for measurement error. A bootstrap algorithm for obtaining confidence intervals for the regression parameter which takes into account the variability due to estimation of the reliability coefficient is also proposed.

METHODS

Measurement error model and correction of logistic regression parameter

The methods described below are based on the measurement error model of Armstrong et al. (7) for matched case-control studies. We assume that in each matching stratum, a case is matched to a variable number of controls. However, the techniques are generalizable to the unmatched design by assuming that there is only one matching stratum.

Let x_{ijk} denote the unobserved true value of the exposure variable for the kth subject with case/control status j (0 = control, 1 = case), in stratum i ($i = 1, \ldots, M$). Assume that x_{ijk} is normally distributed with mean, $\mu_i + j\delta$, and variance σ_s^2. In addition, let z_{ijkl} denote the lth observed value of x_{ijk}, measured with error, for $l = 1, \ldots, n_{ijk}$. We assume the following classical errors-in-variables model:

$$z_{ijkl} = x_{ijk} + e_{ijkl},$$

where the error term, e_{ijkl}, is independent of x_{ijk} and $e_{ijkl'}$, for $l \neq l'$, and normally distributed with mean 0 and variance, σ_e^2. It follows that the observed z_{ijkl} in stratum i are normally distributed with means $\mu_i + \delta$ and μ_i for cases and controls, respectively, and common variance, $\sigma_s^2 + \sigma_e^2$. The variance component, σ_s^2, can be interpreted as the variance of the true exposure, after stratifying by matching stratum and case/control status, and σ_e^2 as the variance due to measurement error.

With these assumptions and the application of Bayes' rule, Armstrong et al. (7) showed that the probability that a study subject is a case, conditional on $\bar{z}_n$, the observed average based on n measurements, and membership in stratum i, is a logistic function:

$$\Pr(D = 1|\bar{z}_n; i) = \frac{\exp(\alpha_i + \beta R_n \bar{z}_n)}{1 + \exp(\alpha_i + \beta R_n \bar{z}_n)}, \quad (1)$$

where

$$R_n = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2/n} \quad (2)$$

is the reliability of $\bar{z}_n$ as a measure of x. When no measurement error is present, $\bar{z}_n = x$, the reliability coefficient is equal to 1, and equation 1 reduces to:

$$\Pr(D = 1|x; i) = \frac{\exp(\alpha_i + \beta x)}{1 + \exp(\alpha_i + \beta x)}.$$

Thus, an estimate of the logistic regression coefficient based on $\bar{z}_n$ will estimate the "naive" coefficient, $\beta^* = \beta R_n$, rather than the true β. Because the reliability coefficient is between 0 and 1, the "naive" β^* will be attenuated relative to β. We can see from equation 2, however, that as the number of repeated measurements increases, the reliability coefficient approaches one, and the corresponding attenuation in β will diminish.

When all subjects have the same number of n repeated measurements, an estimate of the true regression coefficient can be obtained by fitting the logistic model using $\bar{z}_n$ for each subject's exposure measurement, and multiplying the resulting coefficient estimate, $\hat{\beta}^*$, by $1/R_n$. If subjects have a variable number of measurements, however, this approach cannot be applied, since the reliability of the exposure variable is no longer constant for all subjects, but depends on the number of available repeated measurements.

For the case where the reliability of the exposure differs across subjects, a corrected estimate of the regression coefficient may be obtained by multiplying each subject's average exposure measurement by the reliability of the average, prior to model fitting. That is, if the kth subject in stratum i has the observed average $\bar{z}_{ijk\cdot}$, based on n_{ijk} approximate measurements

of x_{ijk}, then replacing the unknown x_{ijk} in the conditional logistic model with the transformed average, $R_{n_{ijk}}\bar{z}_{ijk}.$, where $R_{n_{ijk}}$ is calculated from equation 2, will yield an estimate of the true β. Because the reliability increases with the number of measurements, this transformation results in greater "shrinkage" of averages based on a small number of repeats and less shrinkage of more informative averages based on many repeats.

When all subjects within the same matched set have the same number of repeats, this method is equivalent to the two-stage approach proposed by Thomas et al. (4) and Whittemore (8) for error correction in linear models, in which $E(x_{ijk}|\bar{z}_{ijk}.)$ is computed and then used as the exposure in the usual regression model. Given the model assumptions described above,

$$E(x_{ijk}|\bar{z}_{ijk}.) = R_{n_{ijk}}\bar{z}_{ijk}. + (1 - R_{n_{ijk}})E(x_{ijk}).$$

If n_{ijk} is constant for all subjects in stratum i, the $(1 - R_{n_{ijk}})E(x_{ijk})$ term is absorbed in the intercept term and does not affect the estimate of the slope parameter in the logistic regression model. Thus, utilizing $E(x_{ijk}|\bar{z}_{ijk}.)$ or $R_{n_{ijk}}\bar{z}_{ijk}.$ will yield equivalent estimates of the true regression parameter. Furthermore, when all subjects in the study have the same number of n repeats, this technique will result in a corrected estimate of the logistic regression coefficient that is identical to the one obtained by correcting the naive estimate by $1/R_n$.

Although fitting the logistic model to the transformed covariate will result in an unbiased estimate of β, the corresponding variance of $\hat{\beta}$ will be underestimated unless the variance components in the reliability coefficient are known. Usually, however, the variance components are estimated from a separate reliability substudy or from subjects in the main study. In our setting, repeated measurements are assumed to be available on all or a subset of the main study participants. Thus, we can estimate the variance components, σ_s^2 and σ_e^2, from the main study data by fitting the following mixed effects analysis of variance model to data on all cases and controls:

$$z_{ijkl} = \mu_i + \delta_j + \gamma_{ijk} + \epsilon_{ijkl}, \qquad (3)$$

where z_{ijkl} is defined as before, μ_i is the effect for stratum i, δ_j is the effect due to case/control status, γ_{ijk} is a normally distributed random subject effect with mean 0 and variance σ_s^2, and ϵ_{ijkl} is the residual error which is normally distributed with mean 0 and variance σ_e^2.

The variance components, σ_s^2 and σ_e^2, can be estimated using one of several methods, including traditional analysis of variance (ANOVA), maximum likelihood (ML) or restricted maximum likelihood (REML) methods. The ANOVA method, available in the SAS procedure, PROC GLM (SAS Institute, Cary, North Carolina), was used in our example because it is computationally simpler than the other methods. This is an important consideration when implementing the bootstrap procedure that we describe in the next section for generating confidence intervals. However, the ANOVA method can lead to negative variance estimates. The ML or REML estimators, which are available from PROC MIXED in SAS, do not have this limitation. For further details about the different estimation techniques, see Searle et al. (9).

The steps involved in obtaining an estimate of the logistic regression coefficient corrected for measurement error can be summarized as follows:

1. Estimate the variance components, σ_e^2 and σ_s^2, by fitting the mixed effects analysis of variance model in equation 3 to the study data.
2. Multiply each subject's average exposure by

$$\hat{R}_{n_{ijk}} = \hat{\sigma}_s^2/(\hat{\sigma}_s^2 + \hat{\sigma}_e^2/n_{ijk}).$$

3. Estimate the true logistic regression coefficient, β, by fitting a conditional logistic regression model to the transformed averages.

Because the technique is based on assuming that the true exposure and measurement error are normally distributed, suitable data transformations should be applied when the distributions deviate from normality. Note, however, that a data transformation such as the log-transform will result in a model in which the log odds of disease is a linear function of the exposure measured on the log rather than the original scale.

Bootstrap method for obtaining confidence intervals

The width of the usual 95 percent confidence interval for the true β based on the transformed covariate will be too narrow because the interval does not account for the extra variability associated with estimation of the variance components in R_n. Rosner et al. (2) have derived the asymptotic variance and corresponding confidence intervals of the corrected logistic regression parameter which includes the uncertainty of the variance estimates for use in cohort studies under a rare disease assumption. Their method, however, is applicable only when all subjects in the main study have the same number of repeats. For the situation when subjects in a matched case-control study have a variable number of replicates, we propose the follow-

ing bootstrap algorithm for obtaining confidence intervals for the true β:

1. Assuming there are M matched sets in the main study, generate a bootstrap sample using the matching stratum as the sampling unit, and sampling M matched sets with replacement from the main study data. For each matched set that is selected, the sample contains all the subjects within the matched set, along with each subject's case/control status and repeated measurements.
2. Using the bootstrap sample, estimate σ_s^2, σ_e^2, and the true β by following the three-step approach outlined in the previous section.
3. Repeat steps 1) and 2) 1,000 times, which is the approximate minimum number of bootstraps necessary to compute bias-corrected confidence limits (10).

In constructing a bootstrap sample from the main study data, sampling occurs at the level of the matching stratum because the matching between the cases and the controls needs to be preserved. If the number of controls matched to each case is variable across strata, one can sample the strata according to the number of subjects in each matched set, in order to keep the total sample size constant for each bootstrap iteration. For example, one samples with replacement M_2 matched sets from the M_2 sets in the main study with two controls per case, M_3 sets from the strata with three controls per case, etc.

The simple $(1 - \alpha)\%$ confidence interval can be constructed using the $\alpha/2$ and $(1 - \alpha/2)$ percentiles of the bootstrap distribution. Bias-corrected confidence intervals should be used when the bootstrap distribution of β is asymmetric and when the sample size is small (10). We report only the bias-corrected confidence intervals in this paper.

Thus far, our focus has been on correcting for measurement error in a single exposure variable, in the absence of confounders. However, the methods can also be generalized to the multi-covariate situation, where the confounders, in addition to the primary exposure variable, may be measured with error. A brief outline of the methods is given in the Appendix. Additional details on the measurement error model and estimation of variance components are also described in Armstrong et al. (7).

EXAMPLE

The primary aim of the NYU Women's Health Study is to determine whether serum levels of endogenous hormones, such as estradiol, are associated with risk of breast cancer. Between March 1985 and June 1991, a cohort of 14,275 healthy women aged 34–65 years were enrolled at the Guttman Breast Diagnostic Institute in New York City. At the time of enrollment and at annual screening visits thereafter, women were asked to donate blood and complete a self-administered questionnaire. Serum samples were frozen and stored for future biologic assays. Subsequent cases of breast cancer were identified primarily through active follow-up and confirmed by reviewing medical and pathologic records. In this example, only the women who were postmenopausal at enrollment (49 percent) were included.

In order to limit the costs associated with measuring hormone levels in the cohort, a nested case-control study design was used. For each incident case of breast cancer, individually matched controls were selected at random from the risk set consisting of all cohort members alive and free of breast cancer at the time of diagnosis of the case, and who matched the case on menopausal status at entry, age at entry, and number and approximate dates of blood donations up to the case's date of diagnosis. For additional details of the study design, see Toniolo et al. (5).

The goal of this example is to evaluate the effect of random measurement error on the associations between total estradiol, % free estradiol, and % estradiol bound to sex hormone-binding globulin (SHBG) and risk of breast cancer, when the average of all the available repeated measurements for a subject is used as her exposure. The associations between the baseline measurements of the total estradiol, % free estradiol, and % estradiol bound to SHBG and risk of breast cancer among postmenopausal women, unadjusted for measurement error, were evaluated by Toniolo et al. (11). Total estradiol and % free estradiol were found to be positively associated with risk of breast cancer, whereas % estradiol bound to SHBG had a strong protective effect.

One of the assumptions of the measurement error model is that the true and observed exposure variables are normally distributed. The distribution of total estradiol levels was skewed, so the logarithms of the values were used. Based on data from both postmenopausal cases and controls, we estimated the reliability coefficients for total estradiol, % free estradiol, and % estradiol bound to SHBG, adjusted for matching stratum and case/control status, as 0.48, 0.68, and 0.92, respectively (table 1). (These estimates were somewhat lower than the estimates published by Toniolo et al. (12), 0.51, 0.77, and 0.94 for total estradiol, % free estradiol, and % estradiol bound to SHBG, respectively, based on data from only the postmenopausal controls in the NYU Women's Health Study.) The estimates of the reliability coefficients indicate that the degree of measurement error in total estradiol and %

TABLE 1. Reproducibility of total estradiol, % free estradiol and % estradiol bound to sex hormone-binding globulin (SHBG), adjusted for case/control status and matching stratum in a nested case-control study of serum hormone levels and breast cancer: New York University Women's Health Study (5)

Hormone	Within-subject variance	Between-subject variance	Reliability coefficient
Total estradiol	0.17	0.16	0.48
% free estradiol	0.017	0.036	0.68
% SHBG-bound estradiol	9.38	104.45	0.92

free estradiol may be sufficiently large to attenuate observed relations with risk of breast cancer.

The main case-control study sample consisted of 381 subjects stratified into 130 matched sets. Ten matched sets had one control per case, 119 sets had two controls per case, and one set had three controls per case. Of the 381 subjects in the main study, 157 (41 percent) had two or more repeated measurements, i.e., 98 subjects had two replicates, 53 had three replicates, and six subjects had four replicates.

We investigated the effects of measurement error on the observed associations between each exposure variable and risk of breast cancer by comparing the estimated logistic regression parameters based on the first measurement of the exposure for each subject, the average of the replicate measures, and the transformed (corrected) average value. Corresponding odds ratios were calculated from the regression estimates by comparing women in the 90th versus 10th percentiles of the observed distributions (i.e., 63.0 vs. 14.5 for total estradiol, 1.7 vs. 1.04 for % free estradiol, and 57.6 vs. 27.3 for % estradiol bound to SHBG).

The bootstrap confidence intervals were generated using the SAS macro facility to create the bootstrap sample, in conjunction with PROC PHREG, which fits conditional logistic regression models. All analyses were run on a DEC 3000/700 AXP computer workstation (Digital Equipment Corporation, Maynard, Massachusetts).

The results are shown in table 2. For total estradiol and % free estradiol, the uncorrected analyses show that using the observed average of the repeated measurements results in a minor increase in the regression coefficient estimates compared with using only the baseline measurement. On the other hand, the estimated regression coefficients corrected for measurement error using the transformed averages are substantially larger than the estimates based on the observed averages for both variables: increases are 74 percent and 40 percent for total estradiol and % free estradiol, respectively.

TABLE 2. Corrected and uncorrected logistic regression parameter estimates, confidence intervals, and odds ratios for the associations of total estradiol, % free estradiol and % estradiol bound to sex hormone-binding globulin (SHBG) with risk of breast cancer in a nested case-control study of serum hormonal levels and breast cancer: New York University Women's Health Study (5)

Exposure variable	Regression coefficient	95% confidence interval*	Odds ratio†
Total estradiol‡			
First measurement	0.66	0.24 to 1.09	2.64
Average	0.77	0.32 to 1.22	3.10
Transformed average	1.34	0.61 to 2.47	7.16
% free estradiol			
First measurement	1.70	0.69 to 2.71	3.07
Average	1.73	0.70 to 2.77	3.13
Transformed average	2.42	1.06 to 4.00	4.95
% SHBG-bound estradiol			
First measurement	−0.046	−0.068 to −0.024	0.25
Average	−0.045	−0.067 to −0.023	0.26
Transformed average	−0.048	−0.074 to −0.025	0.24

* 95% confidence intervals using transformed average based on bias-corrected bootstrap estimate.

† Comparing women at 90th vs. 10th percentile of observed distribution.

‡ Total estradiol measurements were log-transformed.

The effect of measurement error on the estimated odds ratios is especially striking. When comparing women in the 90th percentile versus the 10th percentile of the observed total estradiol distribution, the corrected odds ratio was estimated to be 7.16, compared with uncorrected odds ratios of 2.64 and 3.10 using the baseline and untransformed average, respectively. Similarly, the corrected odds ratio for % free estradiol was 4.95, compared with 3.07 for the baseline measurement and 3.13 for the average value.

This illustrates how using the observed average of replicate measurements of exposure for each subject may not be sufficient to offset the effects of measurement error when the degree of error is large and when subjects have only a few replicates, and that additional error correction procedures may be necessary. In the case of total estradiol, one would need to take the average of 10 replicate measurements to improve the reliability to 0.90, based on the estimated variance components in table 1. For % free estradiol, one would need five measurements. Thus, it is not surprising that using the average value in our example did not appreciably deattenuate the corresponding regression coefficient, because only 41 percent of the study subjects had replicate measurements, and, among these, most had only two or three measurements. Using the average resulted in a 17 percent increase in the regression coefficient for total estradiol, relative to using the first measurement. In comparison, if all subjects had two

replicates, the expected increase in the regression coefficient would be

$$(R_2 - R_1)/R_1 = (0.65 - 0.48)/0.48 = 37\%$$

over the estimate based on one measurement. On the other hand, because levels of % estradiol bound to SHBG are highly reproducible, the logistic regression estimates and corresponding odds ratios using the corrected average were not very different from the uncorrected analyses.

Because 119 (92%) of the 130 matched sets had two controls matched per case, implementation of a more complex stratified bootstrap sampling scheme, which would keep the total number of subjects constant for each iteration, was not warranted. As one would expect, the bias-corrected bootstrap confidence intervals based on the transformed average, as shown in table 2, are shifted further away from 0 and are wider than the uncorrected confidence intervals for all variables, because the bootstrap method accounts for the variation due to estimation of the variance components in the reliability coefficient. When the variation in the estimates of variance components estimates was ignored, the simple 95 percent confidence intervals based on the corrected average were estimated to be 0.54 to 2.13, 1.02 to 3.82, and −0.074 to −0.025 for total estradiol, % free estradiol, and % estradiol bound to SHBG, respectively. Thus, ignoring the extra source of variation from $\hat{R}_{n_{ijk}}$ underestimated the width of the confidence interval by as much as 17 percent (for total estradiol) in our data set.

DISCUSSION

In most reliability studies, the within-subject or error variance of the exposure is estimated from an external population or from a random subset of the main study population from whom repeated measurements are obtained, and one must assume that the resulting estimate is generalizable to the main study population. In our example, the within-subject variances were estimated from the subjects in the main study with at least two repeated hormone measurements. Women with repeated measurements, however, may differ from women with only one measurement. Because blood samples in the NYU Women's Health Study were obtained at annual breast cancer screening visits, women with a family history of breast cancer, for example, or those who are more health conscious, may have been more likely to return for subsequent visits. It is unlikely, though, that this would result in a systematic difference in the within-subject variability of the hormone levels between the subset with repeats and those who had only one measurement. Thus, we can assume generalizability of the estimated within-subject variance to all subjects in our main study.

A second assumption of our error-correction method is that a subject's measurements are distributed randomly around the unobserved true value, and that levels of the exposure are not changing systematically over time. This assumption may not be true if hormone levels decrease with age. In addition, for breast cancer cases, hormone levels could be influenced by the development of disease so that measurements obtained closer to the date of diagnosis may exhibit a systematic time trend. Among subjects in the NYU Women's Health Study, however, a trend in estradiol levels over time was not observed in preliminary analyses using linear regression techniques (results not shown).

We have also assumed that the variance components, σ_s^2 and σ_e^2, are homogeneous across strata and case/control status. The within-subject variance for total estradiol was estimated as 0.16 and 0.18 for cases and controls, respectively, indicating that the error variances are similar for the two groups. Because only one case was included in each stratum, we could not evaluate whether σ_s^2 was constant for cases and controls. Furthermore, assessing whether σ_s^2 was homogeneous across strata was not possible, given that most strata only had two controls.

The error-correction methods in this paper are applicable to studies in which a variable number of repeated measurements of exposure are obtained on subjects, and the average of each subject's measurements is used as the exposure variable. In principle, a corrected estimate of the logistic regression coefficient could also be obtained by utilizing only the first measurement of exposure for each subject, and correcting the resulting estimate by the reliability of a single measure. Although this method is much simpler than using all the available repeated measurements and applying the techniques proposed in this paper, the estimate based on a single measurement will not be as efficient. For example, the 95 percent bias-corrected bootstrap confidence interval for the true β using only the first measurements of total estradiol was 0.52–2.73, which is wider by 19 percent than the corresponding interval based on the transformed averages.

Haukka (13) proposed a similar bootstrap method for correcting for measurement error in generalized linear models for the situation when the "gold standard" is known for the exposure measurement and when validation (as opposed to reproducibility) data are available. When compared with the correction method for logistic regression proposed by Rosner et al. (14), which also takes into account the variability in $\hat{R}$, the bootstrap method was found to yield wider

confidence intervals for peaked and skewed measurement error distributions. As discussed by Haukka (13), this difference may result because the bootstrap method takes better account of the measurement error variance, whereas the Rosner et al. method (14) is based on a first-order Taylor series approximation, which may not adequately correct confidence intervals when the error variance is large.

We have shown that in situations when the magnitude of measurement error is large and subjects have only a few repeats, using the average of the available replicate measurements for each subject may not be sufficient to adjust for the measurement error. The methods proposed in this paper can be applied to provide additional correction procedures in the analysis of case-control data where subjects have a variable number of repeated measures of the exposure. The advantage of our algorithm is that it is conceptually straightforward and relatively easy to implement, especially with the amount of computing power that is now readily available to most investigators.

ACKNOWLEDGMENTS

This work was supported by grant no. DAMD17-94-J-4174 from the Department of Defense Breast Cancer Research Program and grant nos. CA-16087 and CA-34588 from the National Institutes of Health.

The authors are grateful to Drs. Alice Whittemore and Stuart Lipsitz for helpful comments on the manuscript.

REFERENCES

1. De Klerk NH, English DR, Armstrong BK. A review of the effects of random measurement error on relative risk estimates in epidemiological studies. Int J Epidemiol 1989;18:705–12.
2. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol 1992;136:1400–13.
3. Armstrong BG. The effects of measurement errors on relative risk regressions. Am J Epidemiol 1990;132:1176–84.
4. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. Annu Rev Public Health 1993;14:69–93.
5. Toniolo P, Pasternack B, Shore R, et al. Endogenous hormones and breast cancer: a prospective cohort study. Breast Cancer Res Treat 1991;18:S23–S26.
6. Liu X, Liang KY. Efficacy of repeated measures in regression models with measurement error. Biometrics 1992;48:645–54.
7. Armstrong B, Whittemore A, Howe G. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. Stat Med 1989;8:1151–63.
8. Whittemore AS. Errors-in-variables regression using Stein estimates. Am Statistician 1989;43:226–8.
9. Searle SR, Casella G, McCulloch CE. Variance components. New York: John Wiley and Sons, 1992.
10. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1:54–77.
11. Toniolo P, Levitz M, Zeleniuch-Jacquotte A, et al. A prospective study of endogenous estrogens and breast cancer in postmenopausal women. J Natl Cancer Inst 1995;87:190–7.
12. Toniolo P, Koenig K, Pasternack B, et al. Reliability of measurements of total, protein bound, and unbound estradiol in serum. Cancer Epidemiol Biomarkers Prev 1994;3:47–70.
13. Haukka JK. Correction for covariate measurement error in generalized linear models—a bootstrap approach. Biometrics 1995;51:1127–32.
14. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol 1990;132:734–45.
15. Kim MY, Pasternack BP, Carroll RJ, et al. Estimating the reliability of an exposure variable in the presence of confounders. Stat Med 1995;14:1437–46.

APPENDIX

In order to generalize the techniques to the multivariate situation, assume that x_{ij} denotes a $(p \times 1)$ vector of true covariates for the jth subject in stratum i, and that it follows a multivariate normal distribution with mean vector $\mu_i + \Delta$ for the cases and μ_i for the controls, and covariance matrix Σ. In addition, let

$$z_{ijk} = x_{ij} + e_{ijk}$$

denote the kth observed measurement of x_{ij}, for $k = 1, \ldots, n_{ij}$, where the e_{ijk} are independent and identically distributed according to a multivariate normal distribution with covariance matrix, Ω.

Under these assumptions, Armstrong et al. (7) showed that the probability that a subject is a case, conditional on the mean of n repeated observations of the covariate vectors, $\bar{z}_n$, is equal to the following logistic function:

$$\Pr(D = 1|\bar{z}_n, i) = \frac{\exp(\alpha_i + \bar{z}_n\Lambda_n\beta)}{1 + \exp(\alpha_i + \bar{z}_n\Lambda_n\beta)},$$

where

$$\bar{z}_n = \Lambda_n = (\Sigma + n^{-1}\Omega)^{-1}\Sigma,$$

and β is the $(p \times 1)$ vector of logistic regression parameters.

When subjects have a variable number of replicate measures of the exposure variables, it follows that as in the single covariate case, one can transform the observed mean covariate vector for each subject by multiplying the vector by an estimate of the matrix, $\Lambda_{n_{ij}}$, and then fitting the usual logistic regression model to the transformed covariates to obtain the

corrected logistic regression coefficients for all covariates. A bootstrap algorithm analogous to that for the single covariate case could be used to obtain corrected confidence intervals which take into account the variation due to estimation of $\Lambda_{n_{ij}}$, but the method could become very computationally intensive with a large number of covariates, since more complicated multivariate analysis of variance (MANOVA) models would be needed to estimate Σ and Ω. For the special case when only a single covariate is measured with error and the others are measured without error, however, estimation of the variance components is greatly simplified (see Kim et al. (15)), and the bootstrap method can be more easily applied.

American Journal of Epidemiology

Vol. 145, No. 11
Printed in U.S.A.

Relation of Serum Levels of Testosterone and Dehydroepiandrosterone Sulfate to Risk of Breast Cancer in Postmenopausal Women

Anne Zeleniuch-Jacquotte,[1] Peter F. Bruning,[2] Johannes M. G. Bonfrer,[2] Karen L. Koenig,[1] Roy E. Shore,[1] Mimi Y. Kim,[1] Bernard S. Pasternack,[1] and Paolo Toniolo[1]

The authors examined the relation between postmenopausal serum levels of testosterone and dehydroepiandrosterone sulfate (DHEAS) and subsequent risk of breast cancer in a case-control study nested within the New York University Women's Health Study cohort. A specific objective of their analysis was to examine whether androgens had an effect on breast cancer risk independent of their effect on the biologic availability of estrogen. A total of 130 cases of breast cancer were diagnosed prior to 1991 in a cohort of 7,054 postmenopausal women who had donated blood and completed questionnaires at a breast cancer screening clinic in New York City between 1985 and 1991. For each case, two controls were selected, matching the case on age at blood donation and length of storage of serum specimens. Biochemical analyses were performed on sera that had been stored at −80°C since sampling. The present report includes a subset of 85 matched sets, for whom at least 6 months had elapsed between blood donation and diagnosis of the case. In univariate analysis, testosterone was positively associated with breast cancer risk (odds ratio (OR) for the highest quartile = 2.7, 95% confidence interval (CI) 1.1–6.8, $p < 0.05$, test for trend). However, after including % estradiol bound to sex hormone-binding globulin (SHBG) and total estradiol in the statistical model, the odds ratios associated with higher levels of testosterone were considerably reduced, and there was no longer a significant trend (OR for the highest quartile = 1.2, 95% CI 0.4–3.5). Conversely, breast cancer risk remained positively associated with total estradiol levels (OR for the highest quartile = 2.9, 95% CI 1.0–8.3) and negatively associated with % estradiol bound to SHBG (OR for the highest quartile = 0.05, 95% CI 0.01–0.19) after adjustment for serum testosterone levels. These results are consistent with the hypothesis that testosterone has an indirect effect on breast cancer risk, via its influence on the amount of bioavailable estrogen. No evidence was found of an association between DHEAS and risk of breast cancer in postmenopausal women. *Am J Epidemiol* 1997;145:1030–8.

androgens; breast neoplasms; DHEAS; estrogens; testosterone

A possible role of androgens in the development of breast cancer in postmenopausal women was first suggested by Grattarola et al. (1). Mechanisms by which androgens may increase breast cancer risk were reviewed by Secreto et al. (2, 3) and Bernstein and Ross (4). Androgens may act directly, by stimulating breast cell proliferation through binding to androgen receptors or by stimulating the synthesis of growth factors inside the breast epithelium. Androgens may also act indirectly through their conversion to estrogens, which are known to stimulate breast cell proliferation (5): aromatization of androstenedione and testosterone in peripheral tissues is the main source of estrogens in postmenopausal women. In addition, it is well established that testosterone binds to sex hormone-binding globulin (SHBG) with greater affinity than estradiol. Testosterone may thus indirectly increase the risk of breast cancer by decreasing the fraction of estradiol bound to SHBG and thereby increasing the nonbound fraction, which is thought to be the fraction available to breast cells (6). Finally, it has been suggested that testosterone inhibits hepatic secretion of SHBG (7), which could also result in a decreased fraction of estradiol bound to SHBG.

Several case-control studies have reported on the association of plasma or serum levels of testosterone with risk of breast cancer in postmenopausal women. Most such studies (2, 8–11, Bruning et al., unpub-

Received for publication July 3, 1996, and accepted for publication January 29, 1997.

Abbreviations: CI, confidence interval; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulfate; FSH, follicle-stimulating hormone; NYU, New York University; SHBG, sex hormone-binding globulin.

[1] Nelson Institute of Environmental Medicine and Kaplan Comprehensive Cancer Center, New York University School of Medicine, NY.

[2] The Netherlands Cancer Institute, Amsterdam, The Netherlands.

Reprint requests to Dr. A. Zeleniuch-Jacquotte, Department of Environmental Medicine, New York University Medical Center, 341 East 25th Street, New York, NY 10010.

lished data) although not all studies (12) observed higher levels of testosterone in cases than in controls. Among the three prospective studies which examined the relation of serum levels of testosterone with risk of postmenopausal breast cancer, one found a significant positive association (13), one found a nonsignificant positive association (14), and the third reported no association (15).

Dehydroepiandrosterone (DHEA) is the androgen produced by the adrenal in largest quantity. The physiologic roles of DHEA and of its sulfate (DHEAS), which is thought to be produced exclusively by the adrenal cortex (16), are unknown. They are considered weak androgens, but also appear to have estrogenic properties (17). It has been proposed that DHEA and DHEAS protect against breast cancer in premenopausal women, but increase breast cancer risk in postmenopausal women (17, 18). These conflicting actions could be reconciled by a recent hypothesis: in premenopausal women, DHEA would have an antiestrogenic effect by binding competitively to estrogen receptors, whereas, in postmenopausal women, DHEA would bind to vacant estrogen receptors and enhance estradiol-like effects, thereby stimulating tumor growth (19).

Results from case-control studies of DHEA and DHEAS conducted in postmenopausal women have been mixed (2, 20–22). The three prospective cohort studies which examined the relation of testosterone with breast cancer risk in postmenopausal women also measured DHEAS; one study (23) reported no association, whereas the two others (13, 24) observed a nonsignificant positive association.

We report here on the relation between postmenopausal serum levels of testosterone and DHEAS and subsequent risk of breast cancer in a case-control study nested within a prospective cohort, the New York University (NYU) Women's Health Study. We previously reported a positive association between postmenopausal serum fractions of bioavailable estrogens and risk of breast cancer in this study population (25). A specific objective of our analysis was to examine whether serum levels of androgens have an effect on breast cancer risk independent of their influence on serum levels and biologic availability of estrogens.

MATERIALS AND METHODS

The NYU Women's Health Study cohort

Between March 1985 and June 1991, the NYU Women's Health Study enrolled a cohort of 14,275 women aged 34–65 years at the Guttman Breast Diagnostic Institute, a breast cancer screening center in New York City. Details concerning subject recruitment have been published elsewhere (25, 26). The current report is limited to the 7,054 cohort members who were postmenopausal at the time of enrollment. Participants were classified as postmenopausal if they reported either: 1) no menstrual cycles during the preceding 6 months, 2) a total bilateral oophorectomy, or 3) a hysterectomy without complete oophorectomy prior to natural menopause and were 52 years of age or older. Cohort members donated 30 mL of blood and completed a self-administered questionnaire at enrollment. Blood was drawn prior to breast examination, between 9:00 A.M. and 3:00 P.M. in nonfasting women. After centrifugation, serum samples were immediately stored at −80°C for subsequent biochemical analyses. Women who had taken hormonal medications in the 6 months preceding their visit were not eligible.

Nested case-control study

Cases of breast adenocarcinoma were identified primarily through active follow-up of the cohort and were confirmed by review of individual clinical and pathology records (25). For each case diagnosed in a woman who was postmenopausal at enrollment, two controls were selected at random from the risk set of women who were alive and free of disease at the time of diagnosis of the case, and who matched the case on age at enrollment (±6 months), date of initial blood donation (±3 months), and menopausal status. As of October 1991, 130 members of the postmenopausal cohort had been identified who had received a diagnosis of breast cancer prior to January 1, 1991. Serum assays of follicle-stimulating hormone (FSH) were conducted to confirm the postmenopausal status of all the cases and their selected controls: three controls, who had reported the absence of menstrual cycles in the 6 months prior to enrollment, had FSH levels below 17.5 IU/liter, which was less than the minimal level compatible with postmenopausal status for our assay. They were nonetheless included in the analysis, since excluding them did not materially affect risk estimates. Estrogen assays (total estradiol, % estradiol free, and % estradiol bound to SHBG) were performed for all matched sets. For logistical reasons, androgen assays were carried out in a subset of 118 matched sets. Excluded from the analyses reported here are 33 matched sets for whom diagnosis of the case occurred 6 months or less after blood donation, six controls who reported treatment with corticosteroids in the 6 months prior to blood donation, and one control whose estrogen assays were done on a different day than the matching case. As a result, 85 cases (83 invasive and 2 noninvasive intraductal) and 163 controls are included in the present report.

Laboratory methods

For androgen assays, serum samples that had not been previously defrosted were shipped in dry ice to the Netherlands Cancer Institute and analyzed in two batches. Samples from a case and her matched controls were always analyzed in the same batch. All assays were performed in duplicate with the laboratory personnel blinded to the case or control status of the samples. Reference sera were included for each assay in several places within each batch.

Total testosterone was measured by a solid-phase radioimmunoassay (Coat-A-Count, Diagnostic Products Corp., Los Angeles, California) not requiring extraction or chromatography. The mean intra-assay coefficient of variation in the range of measurement was 6.2 percent. The inter-assay coefficients of variation were, respectively, 11 percent at 1.67, 2.2 percent at 10.00, and 7.0 percent at 22.05 nmol/liter.

DHEAS was measured directly in diluted serum as we have reported previously (27) using an antiserum against DHEA which showed a 42 percent crossreactivity with DHEAS. As DHEA was present in serum in concentrations at least 10 times lower than DHEAS, it had a negligible influence on the DHEAS values, which were read from a DHEAS standard plot. The mean intra-assay coefficient of variation in the range of measurement was 3.3 percent. The inter-assay coefficients of variation were, respectively, 10 percent at 1.68, 9 percent at 2.97, 10 percent at 5.51, and 8 percent at 15.18 μmol/liter.

Total estradiol was measured by standard radioimmunoassay (Pantex, Inc., Santa Monica, California). Percent estradiol bound to SHBG and % estradiol free were measured with a concanavalin A-sepharose binding and an ultrafiltration method, respectively, as reported previously (25).

Statistical methods

When treated as continuous, total estradiol, testosterone, and DHEAS were $\log_e$-transformed to reduce departures from the normal distribution. The paired t test was used to compare hormone levels of the cases to the mean hormone levels of their matched controls.

To compute odds ratios, hormonal measurements were categorized into quartiles, using the frequency distribution of the cases and the controls combined. Because the androgen assays were performed in two batches, quartile cut-points were calculated separately for each batch. The weighted averages of the cut-points are reported in the tables.

The data were analyzed using conditional logistic regression (28). Odds ratios were computed relative to the lowest quartile. Regression analyses were also performed on the continuous hormonal variables. Likelihood ratio tests were used to assess the statistical significance of overall associations, linear trends, and deviations from linearity. All p values are two-sided.

One objective of the analysis was to examine concurrently the effect of androgens and estrogens. Therefore, we report on the effect of adding androgen variables to models containing estrogen variables, and vice-versa. When adding estrogen variables to models containing androgen variables, % SHBG-bound estrogen was entered first because it was the estrogen variable most strongly associated with breast cancer risk in multivariate models (25).

Hormone levels in this study were assessed from a single blood donation. For some hormones, however, a single measurement may not provide a reliable estimate of a woman's long-term average level, the exposure of interest, because of intrinsic fluctuations in the hormone over time and laboratory measurement error. In addition, different hormones are measured with varying amounts of error. For example, the reliability coefficients of total estradiol, % estradiol bound to SHBG, and DHEAS were estimated to be 0.51, 0.94, and 0.75, respectively, in our study population (29). The reliability of testosterone was not assessed in our study, but estimates from the literature range from 0.74 (30) to 0.88 (31). We were concerned that these differences might distort our results regarding the relative importance of the hormones. We therefore applied the method of Armstrong et al. (32) for correcting logistic regression parameter estimates of continuous variables for measurement error in case-control data. For total estradiol, % estradiol bound to SHBG, and DHEAS, we used within-subject variances which we had previously estimated (29). For testosterone, we used the within-subject variance estimate provided by Hankinson et al. (Susan Hankinson, Harvard University School of Public Health, personal communication, 1996). We assumed that the different hormonal variables had independent measurement errors.

We examined the effect of Quetelet index (weight (kg)/height $(m)^2$) on the androgen-breast cancer associations, because the rate of conversion of androgens to estrogens increases with Quetelet index (33), and because the known positive association of Quetelet index with risk of breast cancer was confirmed in our data (25). The effect of other known risk factors (age at menarche, parity, age at first full-term pregnancy, age at menopause, history of breast cancer in a first-degree relative, history of a benign breast condition, history of total oophorectomy, lifetime months of lactation, and smoking history) on the androgen-breast cancer associations was also examined in multivariate

conditional logistic analyses. The inclusion of covariates other than Quetelet index in the statistical analyses did not materially affect the results and are therefore not presented. In addition, the exclusion from the analysis of the six cases and 13 controls who had a total oophorectomy prior to enrollment in the study had no material impact on the results (data not shown). Results are therefore presented including these patients.

RESULTS

Some characteristics of the study group are given in table 1. The median age at diagnosis of breast cancer was 61.6 years and the median duration between blood donation and diagnosis was 2.7 years (range 0.5–5.5 years). Known breast cancer risk factors had a similar distribution in this group as in the larger group on which estrogen assays were carried out (25). There were no appreciable differences between cases and controls in age at menarche, parity, age at menopause, and history of prior oophorectomy. Delayed first full-term pregnancy, history of breast cancer in at least one first-degree relative 45 years old or younger and history of a benign breast condition were associated with a nonsignificant increase in risk of breast cancer, while a history of breastfeeding was associated with a nonsignificant protective effect. The median weight and median Quetelet index were significantly higher in cases than in controls.

Table 2 shows the geometric mean levels of testosterone and DHEAS for cases and controls. The mean testosterone level was 21 percent higher in cases than in controls ($p < 0.01$) and the mean DHEAS level was 20 percent higher ($p = 0.10$).

TABLE 2. Geometric mean, geometric standard deviation, and range of serum levels of testosterone and dehydroepiandrosterone sulfate (DHEAS) in breast cancer patients diagnosed at least 6 months after blood donation and their individually matched controls, New York University Women's Health Study, 1985–1990

Hormone	Cases ($n = 85$)	Controls ($n = 163$)
Testosterone (nmol/liter)		
Mean** (SD†)	1.05 (1.79)	0.87 (1.89)
Range	0.20–3.96	0.14–5.96
DHEAS (μmol/liter)		
Mean* (SD†)	2.36 (2.37)	1.96 (2.26)
Range	0.22–14.60	0.12–10.43

* $p = 0.10$, paired t test.
** $p < 0.01$, paired t test.
† SD, standard deviation.

Table 3 reports odds ratios for the association between breast cancer and serum levels of testosterone, total estradiol, and % estradiol bound to SHBG. In univariate analyses, odds ratios showed a significant increase ($p = 0.02$, test for trend) in risk of breast cancer with increasing levels of testosterone: the odds ratios (95 percent CIs) for the second, third, and fourth quartiles relative to the lowest quartile, were 2.4 (1.0–5.6), 3.5 (1.4–8.4), and 2.7 (1.1–6.8), respectively. However, adjusting for % SHBG-bound estradiol, which was the estrogen variable most strongly associated with breast cancer risk, reduced the odds ratios and removed the significant trend. The odds ratios (95% CIs) were 1.5 (0.6–3.7), 2.0 (0.7–5.2), and 1.3

TABLE 1. Characteristics of study subjects, New York University Women's Health Study, 1985–1990

Characteristic	Cases ($n = 85$)	Controls ($n = 163$)
Age (years) at blood donation, median (range)	59.2 (48.9–65.4)	59.1 (48.9–64.9)
Age (years) at diagnosis, median (range)	61.6 (52.2–68.6)	
Age (years) at menarche, median (range)	13 (9–16)	13 (10–17)
No. of full-term pregnancies (%)		
0	24.7	23.9
1	17.6	13.5
>1	57.6	62.6
Age (years) at first full-term pregnancy, median (range)	25 (16–41)	24 (16–43)
Ever breastfeeding (%)	20.8	28.2
Age (years) at menopause, median (range)	51.7 (31.6–57.2)	50.9 (24.9–58.6)
Breast cancer in first-degree relative <45 years old (%)	8.2	3.7
Prior benign breast condition (%)	57.7	46.7
Prior bilateral oophorectomy (%)	7.0	8.0
Height (cm), median (range)	162.6 (149.9–177.8)	162.6 (147.3–177.8)
Weight* (kg), median (range)	70.3 (47.6–122.5)	62.6 (45.4–124.7)
Quetelet's index* (kg/m^2), median (range)	26.1 (19.9–43.6)	24.0 (17.7–44.4)

* $p < 0.001$, paired t test.

TABLE 3. Odds ratios (OR) and 95% confidence intervals (CI) for the association between breast cancer risk and serum levels of testosterone, total estradiol, and % estradiol bound to sex hormone-binding globulin (SHBG), New York University Women's Health Study, 1985–1990

Hormonal variable by quartiles	Unadjusted OR†	95% CI	Adjusted OR‡	95% CI	Adjusted OR§	95% CI
Testosterone#						
1	1.0		1.0		1.0	
2	2.4	1.0–5.6	1.5	0.6–3.7	1.4	0.6–3.5
3	3.5	1.4–8.4	2.0	0.7–5.2	1.8	0.7–5.0
4	2.7	1.1–6.8	1.3	0.5–3.7	1.2	0.4–3.5
p for trend	*		NS¶		NS	
Total estradiol††						
1	1.0		1.0		1.0	
2	2.0	0.8–5.3	1.8	0.7–4.8	1.7	0.6–4.7
3	4.3	1.8–10.4	3.6	1.4–9.0	2.6	1.0–6.8
4	3.8	1.5–10.3	2.9	1.0–8.3	1.6	0.5–5.8
p for trend	***		*		NS	
% SHBG-bound estradiol‡‡						
1	1.0		1.0		1.0	
2	0.43	0.19–0.98	0.44	0.19–1.01	0.44	0.19–1.05
3	0.19	0.07–0.49	0.20	0.07–0.56	0.21	0.07–0.59
4	0.05	0.01–0.17	0.05	0.01–0.19	0.05	0.01–0.21
p for trend	***		***		***	

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
† Unadjusted, except for matching factors (age and serum storage time).
‡ For testosterone, odds ratios are adjusted for % SHBG-bound estradiol; for total estradiol and % SHBG-bound estradiol, odds ratios are adjusted for testosterone.
§ Adjusted for other hormonal variables in the table.
¶ NS, not significant.
# The cut-points defining quartiles of testosterone were 0.73, 1.02, and 1.45 nmol/mL.
†† The cut-points defining quartiles of total estradiol were 20, 30, and 45 pg/mL.
‡‡ The cut-points defining quartiles of % SHBG-bound estradiol were 34.4, 43.6, and 51.3%.

(0.5–3.7) for second, third, and fourth quartiles, respectively. Adding total estradiol to the model including testosterone and % SHBG-bound estradiol did not significantly improve the fit of the model, although it further reduced the odds ratios (95 percent CIs) to 1.4 (0.6–3.5), 1.8 (0.7–5.0), and 1.2 (0.4–3.5), respectively. Adding % free estradiol or Quetelet index to the model containing testosterone, estradiol, and % estradiol bound to SHBG did not materially affect the odds ratios (data not shown). A strong positive association between breast cancer risk and increasing levels of total estradiol was also present in univariate analysis. This association remained significant after adjusting for testosterone levels, although the odds ratios and the corresponding *p* values were somewhat reduced. The protective effect associated with increasing percentage of SHBG-bound estradiol was hardly affected by adjustment for testosterone levels. In the model including the three hormonal variables, only % estradiol bound to SHBG remained significant. Analysis on continuous variables showed similar results.

Results of the analyses correcting for measurement error were similar to results of the uncorrected analyses with respect to the relative strength of the associations of the hormonal variables with breast cancer risk: the positive association of testosterone was weakened and no longer significant after adjusting for % SHBG-bound estradiol, whereas the positive association of total estradiol became only marginally significant and the negative association of % SHBG-bound estradiol remained highly significant after adjusting for testosterone. In the model including the three variables, only % SHBG-bound estradiol remained significant.

Table 4 reports odds ratios for the association be-

TABLE 4. Odds ratios (OR) and 95% confidence intervals (CI) for the association between breast cancer risk and serum levels of dehydroepiandrosterone sulfate (DHEAS), New York University Women's Health Study, 1985–1990

Quartiles of DHEAS†	OR‡	95% CI	OR§	95% CI
1	1.0		1.0	
2	0.7	0.3–1.5	0.3	0.1–0.9
3	1.0	0.5–2.1	0.5	0.2–1.3
4	1.6	0.7–3.5	0.9	0.4–2.3
p for trend	NS¶		NS	

† The cut-points defining quartiles of DHEAS were 1.33, 2.38, and 3.58 μmol/liter.
‡ Unadjusted, except for matching factors (age and serum storage time).
§ Adjusted for % estradiol bound to sex hormone-binding globulin and total estradiol.
¶ NS, not significant.

tween breast cancer risk and increasing levels of DHEAS. In unadjusted analyses, although the odds ratio in the highest quartile was slightly elevated (odds ratio = 1.6, 95 percent CI 0.7–3.5), there was no trend of increasing risk of breast cancer with increasing levels of DHEAS. The inclusion of estrogen variables or of Quetelet index did not result in a significant trend. The odds ratios for the association between breast cancer risk and DHEAS are shown adjusting for % SHBG-bound estradiol and total estradiol. Inclusion of DHEAS in models containing estrogen variables did not materially affect the associations between estrogen variables and breast cancer risk (data not shown). Correcting for measurement error in the hormonal variables did not alter the results.

Finally, analyses were conducted using only the 56 matched sets with at least 2 years between blood donation and diagnosis of the case. The results were similar to the results of analyses conducted in the larger group, both for testosterone and DHEAS (data not shown).

Table 5 reports the Spearman correlation coefficients for hormone levels and Quetelet index, by case-control status. Note that testosterone was correlated positively with total estradiol (r_S = 0.23 in cases and 0.27 in controls) and negatively with % estradiol bound to SHBG (r_S = −0.27 in cases and −0.33 in controls).

DISCUSSION

In unadjusted analyses (except for matching variables), we observed a statistically significant trend of increasing risk of breast cancer with increasing serum levels of testosterone in postmenopausal women. Because all cases were diagnosed at least 6 months after blood donation (median 2.7 years) and because a similar trend was observed when the analysis was limited to the two-thirds of the cases diagnosed at least 2 years after blood donation, it seems unlikely that the higher levels of testosterone observed in women who subsequently developed the disease compared with controls resulted from the presence of tumors.

Three previous prospective studies (13–15) have examined the association between serum levels of testosterone and breast cancer risk in postmenopausal women. No association was observed in the Rancho Bernardo, California, study (15), in which the age-adjusted mean testosterone levels were 258 pg/ml in 15 cases diagnosed at least one year after blood donation and 261 pg/ml in 400 noncases. However, results from the two other prospective studies (13, 14) are consistent with ours findings. In the Washington County, Maryland, study (14), serum levels were 11 percent higher in 39 cases (mean 304 pg/ml) than in 155 controls (mean 274 pg/ml), although this difference was not statistically significant. Finally, in 24 cases diagnosed during the first 3.5 years of follow-up of a cohort of 4,040 postmenopausal women from northern Italy, the risk ratios (95 percent CIs) for breast cancer associated with the second and third tertiles of testosterone were 4.8 (0.9–25.1) and 7.0 (1.4–36.4), respectively (p for trend = 0.026) (13).

We recently reported a positive association between bioavailable estrogens and subsequent risk of breast cancer in a slightly larger group of postmenopausal women from the NYU Women's Health Study (25). An objective of the present analysis was to examine whether androgens had an effect on breast cancer risk that was independent of their influence on serum levels and biologic availability of estrogen. Results

TABLE 5. Spearman correlation coefficients for androgen and estrogen levels and Quetelet index, New York University Women's Health Study, 1985–1990

	DHEAS†	Total estradiol	% SHBG†-bound estradiol	% free estradiol	Quetelet index
Controls (n = 163)					
Testosterone	0.35***	0.27***	−0.33***	0.25**	0.30***
DHEAS		0.23**	−0.27***	0.25**	0.05
Total estradiol			−0.48***	0.45***	0.43***
% SHBG-bound estradiol				−0.72***	−0.52***
% free estradiol					0.48***
Cases (n = 85)					
Testosterone	0.38***	0.23*	−0.27*	0.12	0.11
DHEAS		0.28*	−0.37***	0.24*	−0.06
Total estradiol			−0.47***	0.19	0.38***
% SHBG-bound estradiol				−0.56***	−0.42***
% free estradiol					0.29**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.
† DHEAS, dehydroepiandrosterone sulfate; SHBG, sex hormone-binding globulin.

showed that, after including estrogen variables (% SHBG-bound estradiol and total estradiol) in our statistical model, the odds ratios associated with higher levels of testosterone were considerably reduced, and there was no longer a significant trend. A similar result was recently observed in the reanalysis of a population-based case-control study conducted in Sweden (9, 34). Whereas, in univariate analysis, a significant positive association was found between testosterone and breast cancer risk, the association disappeared after controlling for estrone (and androstenedione). On the other hand, Berrino et al. (13) did not observe a reduction of the association between breast cancer risk and levels of testosterone when they adjusted for total estradiol. However, multivariate analysis was hampered by the small sample size of the study (24 cases).

We were concerned about the impact of measurement error in the hormonal variables on our results. It is well known that in the absence of confounders, nondifferential measurement error in an exposure variable will result in an attenuation of the true exposure/disease relation. When several variables are measured with error, however, the associations of these variables with disease in a multivariate model may be weaker or stronger than the true associations (35). It is reassuring that, in our analysis, correcting for measurement error did not affect the relative strength of the associations of the hormonal variables with risk of breast cancer.

Our results are consistent with the hypothesis that testosterone has an indirect effect on breast cancer risk, through its association with estrogen levels. The fact that % SHBG-bound estradiol was the estrogen variable which caused the greatest reduction in the testosterone-breast cancer odds ratios suggests that the effect of testosterone on the bioavailability of estrogens may be more important than its role as a precursor of estrogens. An increase in the serum levels of testosterone could lead to a decrease in % estradiol bound to SHBG, because testosterone binds to SHBG with greater affinity than estradiol. However, the modeling studies performed by Dunn et al. (36) as well as in vitro experiments (37) indicate that higher concentrations of testosterone would be required to observe such an effect. Inhibition of the hepatic secretion of SHBG by testosterone could also result in a decrease in % SHBG-bound estradiol, because small changes in SHBG concentration can produce an important reduction in the percentage of hormone bound to this protein (38). In support of this hypothesis, a moderate negative correlation between testosterone and SHBG has been reported by some studies (7, 39, 40) although not all studies (34).

A limitation of our study is that only total testosterone was measured. The free and albumin-bound hormone fractions might be more relevant biologically because these fractions are thought to diffuse readily into the cells (41). Indeed, with regard to estrogen, the variable most strongly related to risk of breast cancer was the % SHBG-bound estradiol, which had a protective effect. Thus, we cannot exclude the possibility that the free and albumin-bound fractions of testosterone might have an independent effect on breast cancer risk.

The lack of an association between DHEAS and breast cancer observed here is consistent with the results of the previous prospective studies which examined the role of this hormone in postmenopausal women. Barrett-Connor et al. (23) measured DHEAS levels in a cohort of 534 women, 50–79 years old, among whom 21 subsequently developed breast cancer, and reported no difference between cases and non-cases. In a case-control study nested within a cohort of approximately 13,000 female residents of Washington County, Maryland, Gordon et al. (24) reported that serum levels of DHEA were significantly higher in 30 postmenopausal women who developed breast cancer 9 years or more after blood donation than in 59 matched controls. However, no statistically significant difference in DHEAS levels was observed, although serum levels of DHEAS were slightly higher in the women who developed breast cancer than in the controls. Finally, an increase in the odds ratios for breast cancer was observed with increasing serum levels of DHEAS in an Italian cohort study, the Study of Hormones and Diet in the Etiology of Breast Cancer (ORDET Study), but this trend was not statistically significant (13). Overall, there is little epidemiologic evidence that DHEAS plays an important role in breast cancer development in postmenopausal women.

In conclusion, elevated serum levels of testosterone were found to be associated with subsequent risk of breast cancer in postmenopausal women. However, this association was considerably reduced and no longer significant after taking into account the effect of serum estrogen levels on breast cancer risk, suggesting that androgens act through their influence on the availability of estrogens via SHBG binding and/or as precursors of estrogens. There was no evidence in our study that the adrenal androgen DHEAS plays a role in breast cancer development. In light of these results, additional research to identify factors influencing testosterone levels in healthy postmenopausal women would be of interest. Among life-style factors such as smoking, obesity, diet, alcohol consumption, and exercise, only obesity has been found to be mar-

ginally associated with higher levels of testosterone (42, 43).

ACKNOWLEDGMENTS

Supported by grant no. CA-34588, by center grant nos. CA-13343 and CA-16087 from the National Cancer Institute, by grant no. ES-00260 from the National Institute of Environmental Health Sciences, and by grant no. DAMD17-94-J-4174 from the Department of Defense.

The authors thank the staff and management of the Guttman Breast Diagnostic Institute for their collaboration. They also thank Anahita Bassiri, Elizabeth Clancy, Jo-Ann Cutrone, Frances Mastrota, Lynne Quinones, Joan Szymczak, and Cristina Trigo for data collection and management. In addition, they express their gratitude to Dorothe Linders and Tiny Korse of the Netherlands Cancer Institute for their technical assistance.

REFERENCES

1. Grattarola R, Secreto G, Recchione C, et al. Androgens in breast cancer. II. Endometrial adenocarcinoma and breast cancer in married postmenopausal women. Am J Obstet Gynecol 1974;118:173–8.
2. Secreto G, Toniolo P, Berrino F, et al. Serum and urinary androgens and risk of breast cancer in postmenopausal women. Cancer Res 1991;51:2572–6.
3. Secreto G, Zumoff B. Abnormal production of androgens in women with breast cancer. (Review). Anticancer Res 1994; 14:2113–17.
4. Bernstein L, Ross RK. Endogenous hormones and breast cancer risk. Epidemiol Rev 1993;15:48–65.
5. Henderson BE, Ross RK, Pike MC. Hormonal chemoprevention of cancer in women. Science 1993;259:633–8.
6. Siiteri PK, Hammond GL, Nisker JA. Increased availability of serum estrogens in breast cancer: a new hypothesis. In: Pike MC, Siiteri PK, Welsh CW, eds. Banbury report 8. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1981: 87–101.
7. Lonning PE, Helle SI, Johannessen DC, et al. Relations between sex hormones, sex hormone binding globulin, insulin-like growth factor binding protein-1 in postmenopausal breast cancer patients. Clin Endocrinol (Oxf) 1995;42:23–30.
8. McFadyen IJ, Forrest APM, Prescott RJ, et al. Circulating hormone concentrations in women with breast cancer. Lancet 1976;1:1100–2.
9. Adami HO, Johansson EDB, Vegelius J, et al. Serum concentrations of estrone, androstenedione and sex-hormone-binding globulin in postmenopausal women with breast cancer and in age-matched controls. Ups J Med Sci 1979;84:259–74.
10. Secreto G, Recchione C, Cavalleri A, et al. Circulating levels of testosterone, 17β-oestradiol, luteinising hormone and prolactin in postmenopausal breast cancer patients. Br J Cancer 1983;47:269–75.
11. Hill P, Garbaczewski L, Kasumi F. Plasma testosterone and breast cancer. Eur J Cancer Clin Oncol 1985;21:1265–6.
12. Malarkey WB, Schroeder LL, Stevens VC, et al. Twenty-four hour pre-operative endocrine profiles in women with benign and malignant breast disease. Cancer Res 1977;37:4655–9.
13. Berrino F, Muti P, Micheli A, et al. Serum sex hormone levels after menopause and subsequent breast cancer. J Natl Cancer Inst 1996;88:291–6.
14. Wysowski DK, Comstock GW, Helsing KJ, et al. Sex hormone levels in serum in relation to the development of breast cancer. Am J Epidemiol 1987;125:791–9.
15. Garland CF, Friedlander NJ, Barrett-Connor E, et al. Sex hormones and postmenopausal breast cancer: a prospective study in an adult community. Am J Epidemiol 1992;135: 1220–30.
16. Vermeulen A. The androgens. In: Gray CH, James VHT, eds. Hormones in blood, 3rd ed, vol 4. New York: Academic Press, Inc, 1983:349–61.
17. Seymour-Munn K, Adams J. Estrogenic effects of 5-androstene-3 beta, 17 beta-diol at physiological concentrations and its possible implication in the etiology of breast cancer. Endocrinol 1983;112:486–91.
18. Bulbrook RD, Hayward JL, Spicer CC. Relation between urinary androgen and corticosteroid excretion and subsequent breast cancer. Lancet 1971;2:395–7.
19. Ebeling P, Koivisto VA. Physiological importance of dehydroepiandrosterone. Lancet 1994;343:1479–81.
20. Zumoff B, Levin J, Rosenfeld RS, et al. Abnormal 24-hr mean plasma concentrations of dehydroepiandrosterone and dehydroepiandrosterone sulfate in women with primary operable breast cancer. Cancer Res 1981;41:3360–3.
21. Jones DL, James VHT. Determination of dehydroepiandrosterone and dehydroepiandrosterone sulphate in blood and tissue: studies of normal women and women with breast or endometrial cancer. J Steroid Biochem 1987;26:151–9.
22. Bernstein L, Ross RK, Pike MC, et al. Hormone levels in older women: a study of postmenopausal breast cancer patients and healthy population controls. Br J Cancer 1990;61: 298–302.
23. Barrett-Connor E, Friedlander NJ, Khaw KT. Dehydroepiandrosterone sulfate and breast cancer risk. Cancer Res 1990; 50:6571–74.
24. Gordon GB, Bush TL, Helzlouer KJ, et al. Relationship of serum levels of dehydroepiandrosterone and dehydroepiandrosterone sulfate to the risk of developing postmenopausal breast cancer. Cancer Res 1990;50:3859–62.
25. Toniolo P, Levitz M, Zeleniuch-Jacquotte A, et al. A prospective study of endogenous estrogens and breast cancer in postmenopausal women. J Natl Cancer Inst 1995;87:190–7.
26. Toniolo PG, Pasternack BS, Shore RE, et al. Endogenous hormones and breast cancer: a prospective cohort study. Breast Cancer Res Treat 1991;18:S23-S26.
27. Bruning PF, Bonfrer JMG, De Jong-Bakker M. Influence of ACTH on aminogluthimidine induced reduction of plasma steroids in postmenopausal breast cancer. J Steroid Biochem 1984;21:293–8.
28. Breslow NE, Day NE. Statistical methods in cancer research. Vol 1. The analysis of case-control studies. IARC Scientific Publications No. 32. Lyon: International Agency for Research on Cancer, 1980.
29. Toniolo P, Koenig KL, Pasternack BS, et al. Reliability of measurements of total, protein-bound, and unbound estradiol in serum. Cancer Epidemiol Biomarkers Prev 1994;3:47–50.
30. Micheli A, Muti P, Pisani P, et al. Repeated serum and urinary androgen measurements in premenopausal and postmenopausal women. J Clin Epidemiol 1991;44:1055–61.
31. Hankinson SE, Manson JE, Spiegelman D, et al. Reproducibility of plasma hormone levels in postmenopausal women over a 2–3 year period. Cancer Epidemiol Biomarkers Prev 1995;4:649–54.
32. Armstrong BG, Whittemore AS, Howe GR. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. Stat Med 1989;8:1151–63.
33. Siiteri PK, MacDonald PC. Role of extraglandular estrogen in human endocrinology. In: Geiger SR, Astwood EB, Greep RO, eds. Handbook of physiology. Washington, DC: American Physiological Society, 1973:615–29.
34. Lipworth L, Adami HO, Trichopoulos D, et al. Serum steroid hormone levels, sex hormone-binding globulin and body mass index in the etiology of postmenopausal breast cancer. Epidemiology 1996;7:96–100.

35. Armstrong BG. The effects of measurement errors on relative risk regressions. Am J Epidemiol 1990;132:1176–84.
36. Dunn JF, Nisula BC, Rodbard D. Transport of steroid hormones: binding of 21 endogenous steroids to both testosterone-binding globulin and cortico-steroid binding globulin in human plasma. J Clin Endocrinol Metab 1981;53:58–68.
37. Bonfrer JMG, Bruning PF, Nooijen WJ. A simple method for the measurement of the steroid fraction bound to sex-hormone bound globulin in serum. J Steroid Biochem 1989;33:227–31.
38. Selby C. Sex hormone binding globulin: origin, function and clinical significance. Ann Clin Biochem 1990;27:532–41.
39. Haffner SM, Newcomb PA, Marcus PM, et al. Relation of sex hormones and dehydroepiandrosterone sulfate (DHEA-SO_4) to cardiovascular risk factors in postmenopausal women. Am J Epidemiol 1995;142:925–34.
40. Maggino T, Pirrone F, Velluti F, et al. The role of the endocrine factors and obesity in hormone-dependent gynecological neoplasias. Eur J Gynaecol Oncol 1993;14:119–26.
41. Pardridge WM. Transport of protein-bound hormones into tissues in vivo. Endocr Rev 1981;2:103–21.
42. Cauley JA, Gutai JP, Kuller LH, et al. The epidemiology of serum sex hormones in postmenopausal women. Am J Epidemiol 1989;129:1120–31.
43. Newcomb PA, Klein R, Klein BEK, et al. Association of dietary and life-style factors with sex hormones in postmenopausal women. Epidemiology 1995;6:318–21.